

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:24:48 ; Search time 32.82 seconds
(without alignments)
22.343 Million cell updates/sec

Title: US-09-765-739A-2

Perfect score: 110

Sequence: 1 NNTTGVFLKQDWGATIKD 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	51	46.4	1191	1 NKCL_SQUAC	P50133 squalus aca
2	47	42.7	540	1 YHBM_YEAST	Q03263 saccharomyc
3	45	40.9	756	1 K6PF_CAEEL	Q27483 caenorhabdi
4	44	40.0	284	1 YQAK_BACSU	P45908 bacillus su
5	44	40.0	465	1 YHJA_ECOLI	P37197 escherichia
6	44	40.0	648	1 VP3_BPPHG	P11129 bacterioph
7	44	40.0	1004	1 ATNA_ARTSF	P28774 artemia san
8	43.5	39.5	290	1 AROE_SYNT3	P74591 synecocyst
9	43	39.1	427	1 PYRC_BACCL	P46538 bacillus ca
10	43	39.1	1021	1 ATN1_CANFA	P50997 canis fami
11	42	38.2	287	1 YUJW_ECOLI	P39409 escherichia
12	42	38.2	307	1 YM16_MYCLE	Q32960 mycobacteri
13	42	38.2	656	1 DNAA_STRCO	P27902 streptomyce
14	42	38.2	1020	1 ATN2_HUMAN	P50993 homo sapien
15	41.5	37.7	366	1 GCST_NEIMA	Q9K018 neisseria m
16	41.5	37.7	368	1 GCST_NEIMA	Q9JVP2 neisseria m
17	41	37.3	377	1 YAE7_METH	Q37139 methanobact
18	41	37.3	395	1 TIME_RAT	P17256 rattus norv
19	41	37.3	749	1 V44_ROTGA	Q04916 rotavirus (
20	41	37.3	866	1 YCBS_ECOLI	P75857 escherichia
21	40.5	36.8	196	1 AFN1_CHICK	P79775 gallus gall
22	40.5	36.8	734	1 PURL_ZYMO	Q9REQ6 zymomonas m
23	40.5	36.8	1025	1 DPYD_HUMAN	Q12882 homo sapien
24	40	36.4	84	1 GVM1_HALN1	P24377 halobacteri
25	40	36.4	120	1 YRAJ_BACSU	Q07934 bacillus su
26	40	36.4	191	1 PGHD_URSAR	Q29562 ursus arcto
27	40	36.4	282	1 BIOB_HELPJ	Q9ZJK8 helicobacte
28	40	36.4	282	1 BIOB_HELPJ	Q25956 helicobacte
29	40	36.4	341	1 Y33B_MYCPN	P75302 mycoplasma
30	40	36.4	489	1 UBPT_CAEEL	Q17361 caenorhabdi
31	40	36.4	500	1 AMPA_BACSU	Q32106 bacillus su
32	40	36.4	942	1 ENV_CAEVG	P31627 caprine art
33	40	36.4	1020	1 ATN2_RAT	P06686 rattus norv

RESULT 1

ID	NKCL_SQUAC	STANDARD	PRT	1191 AA
AC	P55013			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	BUMETANIDE-SENSITIVE SODIUM-(POTASSIUM)-CHLORIDE COTRANSPORTER 1			
DE	(NA-K-CL SYMPORTER) (NKCC).			
GN	SLC12A2 OR NKCC1			
OS	Squalus acanthias (Spiny dogfish).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;			
OC	Elasmobranchii; Squala; Squaloidei; Squalidae; Squalus.			
OX	NCBI_TaxID=7797;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PHOSPHORYLATION SITES.			
RC	TISSUE-Rectal gland.			
RX	MEDLINE=94181560; PubMed=8134373;			
RA	Xu J.-C., Lytle C., Zhu T.T., Payne J.A., Benz E. Jr., Forbush B. III;			
RT	"Molecular cloning and functional expression of the bumetanide-sensitive Na-K-Cl cotransporter";			
RL	Proc. Natl. Acad. Sci. U.S.A. 91:2201-2205(1994).			
CC	-!- FUNCTION: ELECTRICALLY SILENT TRANSPORTER SYSTEM WHICH IS A MEDIATOR OF SODIUM AND CHLORIDE REABSORPTION. PLAYS A VITAL ROLE IN THE REGULATION OF IONIC BALANCE AND CELL VOLUME.			
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	-!- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES.			
CC	-!- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; U05958; AAB60617.1; .			
DR	InterPro; IPR002293; AA_rel_permease.1.			
DR	InterPro; IPR002027; Amino.acid_permease.			
DR	InterPro; IPR002443; NaKCl_transportr.			
DR	Pfam; PF00324; aa_permeases; 1.			
DR	PRINTS; PR01207; NaKCLTRNSPRT.			
DR	PRINTS; PR01208; NaKCLTRNSPRT1.			
KW	Transport; Transmembrane; Glycoprotein; Phosphorylation.			
FT	DOMAIN 1 257 CYTOPLASMIC (POTENTIAL).			
FT	TRANSMEM 258 278 POTENTIAL.			
FT	DOMAIN 279 281 EXTRACELLULAR (POTENTIAL).			
FT	TRANSMEM 282 302 POTENTIAL.			
FT	DOMAIN 303 339 CYTOPLASMIC (POTENTIAL).			
FT	TRANSMEM 340 360 POTENTIAL.			
FT	DOMAIN 361 382 EXTRACELLULAR (POTENTIAL).			
FT	TRANSMEM 383 403 POTENTIAL.			
FT	DOMAIN 404 407 CYTOPLASMIC (POTENTIAL).			
FT	TRANSMEM 408 428 POTENTIAL.			
FT	DOMAIN 429 458 EXTRACELLULAR (POTENTIAL).			

P05023 homo sapien
P13607 drosophila
P55012 mus musculus
P55011 homo sapien
P13508 caenorhabdi
P14585 caenorhabdi
P24620 micromonospor
P40756 rana catesbe
P24021 aspergillus
P57922 pasteurella
P81989 cellulomona
Q02278 mycobacteri

ALIGNMENTS

TRANSMEM 459 479 POTENTIAL.
 DOMAIN 480 496 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 497 517 POTENTIAL.
 DOMAIN 518 564 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 565 585 POTENTIAL.
 DOMAIN 586 630 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 631 651 POTENTIAL.
 TRANSMEM 652 672 POTENTIAL.
 DOMAIN 673 689 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 690 710 POTENTIAL.
 DOMAIN 711 775 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 776 796 POTENTIAL.
 DOMAIN 797 1191 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 27 30 POLY-PRO.
 DOMAIN 51 57 POLY-GLY.
 TRANSMEM 67 70 POLY-ALA.
 DOMAIN 80 92 POLY-ALA.
 TRANSMEM 96 99 POLY-ALA.
 DOMAIN 126 129 POLY-PRO.
 TRANSMEM 523 523 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 535 535 N-LINKED (GLCNAC. . .) (POTENTIAL).
 TRANSMEM 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CARBOHYD 189 189 PHOSPHORYLATION.
 TRANSMEM 1114 1114 PHOSPHORYLATION.
 MOD_RES 1191 AA; 129774 MW; BA7BB9815431500C CRC64;
 SQ SEQUENCE 1191 AA; 129774 MW; BA7BB9815431500C CRC64;

Query Match 46.4%; Score 51; DB 1; Length 1191;
 Best Local Similarity 58.8%; Pred. No. 2;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 TGVFGLKQDWDGATIKD 20
 DB 856 TLVFGFKKQDWRQALMKD 872

RESULT 2

ID YMBM_YEAST STANDARD; PRT: 540 AA.
 AC Q03263;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE HYPOTHETICAL 59.6 KDA PROTEIN IN DSK2-CAT8 INTERGENIC REGION.
 GN YMR279C OR YMR021.05C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -|- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN AS THE DRUG RESISTANCE TRANSLUCASE FAMILY). DHA12 SUBFAMILY.
 CC STRONG, TO YEAST ATR1.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; 249704; CAA89777.1;
 DR SGD; S0004892; YMR279C.
 DR Hypothetical protein; Transmembrane.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 62 82 POTENTIAL.
 FT TRANSMEM 109 129 POTENTIAL.
 FT TRANSMEM 132 152 POTENTIAL.

TRANSMEM 170 190 POTENTIAL.
 FT TRANSMEM 204 224 POTENTIAL.
 FT TRANSMEM 233 253 POTENTIAL.
 FT TRANSMEM 273 293 POTENTIAL.
 FT TRANSMEM 316 336 POTENTIAL.
 FT TRANSMEM 335 355 POTENTIAL.
 FT TRANSMEM 373 393 POTENTIAL.
 FT TRANSMEM 399 419 POTENTIAL.
 FT TRANSMEM 430 450 POTENTIAL.
 FT TRANSMEM 462 482 POTENTIAL.
 FT TRANSMEM 503 523 POTENTIAL.
 SQ SEQUENCE 540 AA; 59561 MW; 687D06CB0D70AF91 CRC64;

Query Match 42.7%; Score 47; DB 1; Length 540;
 Best Local Similarity 38.9%; Pred. No. 3.9;
 Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 NVTTCVFLKQDWDGATI 18
 DB 256 NVPTNIHGLSMDWTGSAL 273

RESULT 3

ID K6PF_CAEEL STANDARD; PRT: 756 AA.
 AC Q27483;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE 6-PHOSPHOFRUCTOKINASE (BC 2.7.1.11) (PHOSPHOFRUCTOKINASE)
 DE (PHOSPHOHEXOKINASE).
 GN C50F4.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA McMurray A.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -|- CATALYTIC ACTIVITY: ATP + D-FRUCTOSE 6-PHOSPHATE = ADP + D-FRUCTOSE 1,6-BISPHOSPHATE.
 CC -|- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
 CC -|- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY. TWO DOMAIN SUBFAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; 270750; CAA94737.1;
 DR HSSP; P00512; 4PFK.
 DR WormPep; C50F4.2; CE05467.
 DR InterPro; IPR000023; Phosphofructokinase.
 DR Pfam; PF00365; PFK; 2.
 DR PRINTS; PR00476; PHFRCTKINASE.
 DR PRODOM; PD000707; Phosphofructokinase; 2.
 DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.
 KW Kinase; Transferase; Glycolysis; Repeat.
 SQ SEQUENCE 756 AA; 83301 MW; 26A89B801D286534 CRC64;

Query Match 40.9%; Score 45; DB 1; Length 756;
 Best Local Similarity 53.3%; Pred. No. 12;
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 6 VFGLKQDWDGATIKD 20

```
Db 436 VIGIKHWDGLKND 450
| | | | |
RESULT 4
YQAK_BACSU
ID YQAK_BACSU STANDARD; PRT; 284 AA.
AC P45908;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 32.2 KDA PROTEIN IN SPOILIC-CWLA INTERGENIC REGION.
GN YQAK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=95219086; PubMed=7704261;
RA Takemaru K.-I., Mizuno M., Sato T., Takeuchi M., Kobayashi Y.;
RT "Complete nucleotide sequence of a skin element excised by DNA
rearrangement during sporulation in Bacillus subtilis.";
RL Microbiology 141:323-327(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S., Sato T.,
RA Takeuchi M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP IDENTIFICATION.
RX MEDLINE=96084975; PubMed=7489895;
RA Medigue C., Mosser I., Viari A., Danchin A.;
RT "Analysis of a Bacillus subtilis genome fragment using a co-operative
computer system prototype.";
RL Gene 165:GC37-GC51(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D32216; BAA06925.1; -
DR EMBL; D84432; BAA12386.1; -
DR EMBL; Z99117; CAB14569.1; -
DR Subtilist; BG11262; yqak.
KW Hypothetical protein; Complete proteome
SQ SEQUENCE 284 AA; 32170 MW; F255261D4692ADB7 CRC64;

Query Match 40.0%; Score 44; DB 1; Length 284;
Best Local Similarity 53.8%; Pred. No. 6.1;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 7 FGLKQDWGATIK 19
| | | | |
Db 193 FGWRNDWDAMALK 205

RESULT 5
YHJA_ECOLI
ID YHJA_ECOLI STANDARD; PRT; 465 AA.
AC P37197;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE CYTOCHROME C PEROXIDASE (EC 1.1.1.5).
GN YHJA OR B3518.

Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
region from 76.0 to 81.5 minutes.";
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -!- CATALYTIC ACTIVITY: 2 FERROCYTOCHROME C + H(2)O(2) =
- 2 FERRICYTOCHROME C + 2 H(2)O.
CC -!- PTM: BINDS 3 HEMES (POTENTIAL).
CC -!- SIMILARITY: HIGH, TO P.AERUGINOSA CYTOCHROME C551 PEROXIDASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00039; AAB18494.1; -
DR EMBL; AE000428; AAC76543.1; -
DR EcoGene; EG12244; vjhA.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR003088; Cyt_C1.
DR Pfam; PF00034; cytochrome_C; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 3.
KW Hypothetical protein; Oxidoreductase; Peroxidase; Heme;
KW Electron transport; Complete proteome.
FT BINDING 59 59 HEME 1 (COVALENT) (BY SIMILARITY).
FT BINDING 62 62 HEME 1 (COVALENT) (BY SIMILARITY).
FT METAL 63 63 IRON 1 (HEME PROXIMAL) (BY SIMILARITY).
FT BINDING 207 207 HEME 2 (COVALENT) (BY SIMILARITY).
FT BINDING 210 210 HEME 2 (COVALENT) (BY SIMILARITY).
FT METAL 211 211 IRON 2 (HEME PROXIMAL) (BY SIMILARITY).
FT BINDING 351 351 HEME 3 (COVALENT) (BY SIMILARITY).
FT BINDING 354 354 HEME 3 (COVALENT) (BY SIMILARITY).
FT METAL 355 355 IRON 3 (HEME PROXIMAL) (BY SIMILARITY).
FT METAL 429 429 IRON 3 (HEME DISTAL) (BY SIMILARITY).
SQ SEQUENCE 465 AA; 51570 MW; 9F494A698949E6DA CRC64;

Query Match 40.0%; Score 44; DB 1; Length 465;
Best Local Similarity 52.9%; Pred. No. 10;
Matches 9; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

Qy 6 VFGLKQDWG--ATIKD 20
| | : | | | | | | | |
Db 243 VFNVQFWDGRAATLQD 259

RESULT 6
VP3_BPPH6
ID VP3_BPPH6 STANDARD; PRT; 648 AA.
AC P11129;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last annotation update)
DE P3 PROTEIN.
GN P3.
OS Bacteriophage phi-6.
OC Viruses; dsRNA viruses; Cystoviridae; Cystovirus.
OX NCBI_TaxID=10879;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-13.
RX MEDLINE=88160044; PubMed=3347997;
RA Gottlieb P., Metzger S., Romantschuk M., Carton J., Strassman J.,
RA Bamford D.H., Kalkkinen N., Mindich L.;
```

RT "Nucleotide sequence of the middle dsRNA segment of bacteriophage phi
 RL 6: placement of the genes of membrane-associated proteins.";
 CC Virology 183:183-190(1988).
 CC -!- FUNCTION: P3 PROTEIN IS NECESSARY FOR ADSORPTION ONTO HOST
 CC CELLS.

CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: M17452; AAA68485.1;
 CC PIR: C28648; P3BPF6.
 CC Envelope protein.
 CC SEQUENCE 648 AA; 69178 MW; B188DFE02ACC54E3 CRC64;

Query Match 40.0%; Score 44; DB 1; Length 648;
 Best Local Similarity 41.2%; Pred. No. 15;
 Matches 7; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 6 VEG--LKQWDGATIKD 20
 Db 181 IFGVYKMDWEGSAVD 197

RESULT 7
 ATNA_ARTSF STANDARD; PRT; 1004 AA.
 ID P28774;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DE SODIUM/POTASSIUM-TRANSPORTING ATPASE ALPHA CHAIN (EC 3.6.3.9) (SODIUM
 DE PUMP) (NA+/K+ ATPASE).
 OS Artemia franciscana (Brine shrimp) (Artemia franciscana).
 CC Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda; Anostraca;
 CC Artemiidae; Artemia.
 CC NCBI_TaxID=6661;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92039032; PubMed=1657719;
 RA Macias M.T., Martinez J.L., Palmero I., Sastre L.;
 RT "Cloning of a cDNA encoding an Artemia franciscana Na/K ATPase alpha-
 RT subunit".
 RL Gene 105:197-204(1991).

CC -!- FUNCTION: THIS IS THE CATALYTIC COMPONENT OF THE ACTIVE ENZYME,
 CC WHICH CATALYZES THE HYDROLYSIS OF ATP COUPLED WITH THE EXCHANGE OF
 CC NA & K IONS ACROSS THE PLASMA MEMBRANE. THIS ACTION CREATES THE
 CC ELECTROCHEMICAL GRADIENT OF NA & K, PROVIDING THE ENERGY FOR
 CC ACTIVE TRANSPORT OF VARIOUS NUTRIENTS.
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(IN) + NA(+)(OUT) = ADP +
 CC PHOSPHATE + H(+)(OUT) + NA(+)(IN).
 CC -!- SUBUNIT: COMPOSED OF THREE SUBUNITS: ALPHA (CATALYTIC), BETA
 CC AND GAMMA.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES).

CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X56650; CAA39972.1;
 CC PIR: JH0470; JH0470.
 CC InterPro: IPR001757; E1-E2_ATPase.

DR InterPro: IPR001454; Hydrolase.
 DR InterPro: IPR000661; Na_H_K_ATPase.
 DR Pfam: PF00122; E1-E2_ATPase; 1.
 DR Pfam: PF00702; Hydrolase; 1.
 DR Pfam: PF00689; Na_K_ATPase_C; 1.
 DR Pfam: PF00690; Na_K_ATPase_N; 1.
 DR PRINTS: PR00119; CATATPASE.
 DR PRINTS: PR00121; NAKATPASE.
 DR PROSITE: PS00154; ATPASE_E1_E2; 1.
 KW Hydrolase; Sodium/potassium transport; Transmembrane;
 KW phosphorylation; Atp-binding.
 FT TRANSMEM 76 96
 FT TRANSMEM 110 126 BY SIMILARITY.
 FT TRANSMEM 272 294 BY SIMILARITY.
 FT TRANSMEM 301 329 BY SIMILARITY.
 FT TRANSMEM 768 791 BY SIMILARITY.
 FT TRANSMEM 828 855 BY SIMILARITY.
 FT TRANSMEM 897 918 BY SIMILARITY.
 FT TRANSMEM 934 959 BY SIMILARITY.
 FT MOD_RES 357 357 PHOSPHORYLATION (PROBABLE).
 FT BINDING 489 489 ATP (BY SIMILARITY).
 SQ SEQUENCE 1004 AA; 110699 MW; CE4E6BECE19A78C7 CRC64;

Query Match 40.0%; Score 44; DB 1; Length 1004;
 Best Local Similarity 40.0%; Pred. No. 24;
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 6 VFLKQWDGATIKD 20
 Db 864 LFGLRKHWDRAVND 878

RESULT 8
 AROE_SYNY3
 ID AROE_SYNY3 STANDARD; PRT; 290 AA.
 AC P74591;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SHIKIMATE 5-DEHYDROGENASE (EC 1.1.1.25).
 DE AROE OR SLR1559.
 GN Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hiroseawa M., Sugliura M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -!- CATALYTIC ACTIVITY: SHIKIMATE + NADP(+) = 5-DEHYDROSHIKIMATE +
 CC NADPH.
 CC -!- PATHWAY: FOURTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN THE
 CC BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).
 CC -!- SIMILARITY: BELONGS TO THE SHIKIMATE DEHYDROGENASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: D90916; BAA18699.1;
 CC InterPro: IPR002907; Shikimate_DH.

DR Pfam; PF01488; Shikimate_DH; 1.
 KW Aromatic amino acid biosynthesis; Oxidoreductase; NADP;
 SQ SEQUENCE 290 AA; 31099 MW; 8A2D38EE5D57B303 CRC64;

Query Match 39.5%; Score 43.5; DB 1; Length 290;
 Best Local Similarity 47.8%; Pred. No. 7.6;
 Matches 11; Conservative 0; Mismatches 5; Indels 7; Gaps 1;
 QY 2 TTTGVFG-----LKQWDGAT 17
 Db 105 TTTDVEGFLAPLLELKQWDSGRT 127

RESULT 9
 ID PYRC_BACCL STANDARD; PRT; 427 AA.

AC P46538;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DIHYDROOATASE (EC 3.5.2.3) (DHOASE).
 GN PYRC.
 OS Bacillus caldolyticus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1394;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=DSM 405;
 RX MEDLINE=94282293; PubMed=7516791;
 RA Ghim S.Y., Nielsen P., Neuhaard J.;
 RT "Molecular characterization of pyrimidine biosynthesis genes from the thermophile Bacillus caldolyticus.";
 RL Microbiology 140:479-491(1994).
 CC -!- CATALYTIC ACTIVITY: DIHYDROOATASE + H(2)O = N-CARBAMOYL-L-ASPARTATE.
 CC -!- COFACTOR: THIS ENZYME TIGHTLY BINDS ONE ZINC ATOM PER CHAIN WHICH IS REQUIRED FOR THE CATALYTIC MECHANISM. IT ALSO BINDS WEAKLY TO TWO OTHERS ZINCS WHICH ARE NOT ESSENTIAL FOR ACTIVITY (BY SIMILARITY).
 CC -!- PATHWAY: THIRD STEP IN PYRIMIDINE BIOSYNTHESIS.
 CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE DHOASE FAMILY. SUBFAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X73308; CAA51737.1; -.
 DR MEROPS; M38.972; -.
 DR InterPro; IPR002195; Dihydrooatase.
 DR Pfam; PF00744; Dihydrooatase; 1.
 DR PROSITE; PS00482; DIHYDROOATASE_1; 1.
 DR PROSITE; PS00483; DIHYDROOATASE_2; 1.
 KW Pyrimidine biosynthesis; Hydrolase; Zinc.
 FT METAL 60 60 ZINC (POTENTIAL).
 FT METAL 62 62 ZINC (POTENTIAL).
 SQ SEQUENCE 427 AA; 46047 MW; 759A2AA99F733F4E CRC64;

Query Match 39.1%; Score 43; DB 1; Length 427;
 Best Local Similarity 68.8%; Pred. No. 14;
 Matches 11; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 4 TGVFGLKQWDGATIK 19
 Db 342 TGVFTLKLQVDWLTIK 357

RESULT 10
 ID ATNL_CANFA STANDARD; PRT; 1021 AA.
 AC P50997;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SODIUM/POTASSIUM-TRANSPORTING ATPASE ALPHA-1 CHAIN PRECURSOR
 DE (EC 3.6.3.9) (SODIUM PUMP) (NA+/K+ ATPASE).
 GN ATP1A1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Xie Z., Li H., Liu G., Wang Y., Askari A., Mercer R.W.;
 RT "Cloning of the dog Na/K-ATPase alpha 1 subunit.";
 RL (In) Bamberg E., Schonher W. (eds.);
 RL The sodium pump, pp.49-52, Springer-Verlag, New York (1994).
 RN [2]
 RP SEQUENCE OF 92-307 FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=92258376; PubMed=1316269;
 RA Canessa C.M., Horisberger J.D., Louvard D., Rossier B.C.;
 RT "Mutation of a cysteine in the first transmembrane segment of Na,K-ATPase alpha subunit confers ouabain resistance.";
 RL EMBO J. 11:1681-1687(1992).
 CC -!- FUNCTION: THIS IS THE CATALYTIC COMPONENT OF THE ACTIVE ENZYME, WHICH CATALYZES THE HYDROLYSIS OF ATP COUPLED WITH THE EXCHANGE OF NA & K IONS ACROSS THE PLASMA MEMBRANE. THIS ACTION CREATES THE ELECTROCHEMICAL GRADIENT OF NA & K, PROVIDING THE ENERGY FOR ACTIVE TRANSPORT OF VARIOUS NUTRIENTS.
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (IN) + NA(+) (OUT) = ADP + PHOSPHATE + H(+) (OUT) + NA(+) (IN).
 CC -!- SUBUNIT: COMPOSED OF THREE SUBUNITS: ALPHA (CATALYTIC), BETA AND GAMMA.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2 ATPASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L42173; AAA67372.1; -.
 DR EMBL; X66174; CAA46950.1; -.
 DR InterPro; IPR001757; E1-E2-ATPase.
 DR InterPro; IPR001454; Hydrolase.
 DR InterPro; IPR000661; Na_H_K-ATPase.
 DR Pfam; PF00122; E1-E2-ATPase; 1.
 DR Pfam; PF00702; Hydrolase; 1.
 DR Pfam; PF00689; Na_K-ATPase-C; 1.
 DR Pfam; PF00690; Na_K-ATPase-N; 1.
 DR PRINTS; PR00119; KATATPASE.
 DR PRINTS; PR00121; NAKATPASE.
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 KW Hydrolase; Sodium/potassium transport; Transmembrane;
 KW Multigene family; Phosphorylation; ATP-binding.
 FT PROPEP 1 5
 FT CHAIN 6 1021
 FT CHAIN 6 1021 SODIUM/POTASSIUM-TRANSPORTING ATPASE ALPHA-1 CHAIN.
 FT TRANSMEM 94 115 POTENTIAL.
 FT TRANSMEM 128 148 POTENTIAL.
 FT TRANSMEM 289 311 POTENTIAL.
 FT TRANSMEM 318 346 POTENTIAL.
 FT TRANSMEM 785 808 POTENTIAL.

```

FT TRANSMEM 847 872 POTENTIAL.
FT TRANSMEM 914 934 POTENTIAL.
FT TRANSMEM 951 976 POTENTIAL.
FT MOD_RES 374 374 PHOSPHORYLATION (PROBABLE).
FT BINDING 506 506 ATP (BY SIMILARITY).
FT CONFLICT 109 109 C -> Y (IN REF. 2).
FT CONFLICT 185 185 I -> V (IN REF. 2).
FT CONFLICT 249 249 K -> E (IN REF. 2).
FT CONFLICT 309 309 L -> F (IN REF. 2).
SQ SEQUENCE 1021 AA; 112666 MW; 938A19AA487CBEAA CRC64;

Query Match 39.1%; Score 43; DB 1; Length 1021;
Best Local Similarity 47.1%; Pred No. 36;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 TVGFLKQDWGATIKD 20
   : ||: ||| | |
Db 879 THLLGLRVMDRWDIND 895

RESULT 11
YJWV_ECOLI
ID YJWV_ECOLI STANDARD; PRT; 287 AA.
AC P39409;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 31.5 KDA PROTEIN IN OSMY-DEOC INTERGENIC REGION.
GN YJWV OR B4379.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.,
RA "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes."
RL Nucleic Acids Res. 23:2105-2119(1995).
CC -1- COFACTOR: BINDS A 4FE-4S CLUSTER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ORGANIC RADICAL ACTIVATING ENZYMES
CC FAMILY.
CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14003; AAA97275.1; .
CC EMBL; AE000508; AAC77332.1; .
CC HSSP; P00198; lFDN.
CC EcoGene; EG12599; YJWV.
CC InterPro; IPR001450; 4FE4S_ferrixdin.
CC InterPro; IPR001989; Radical_activat.
CC Pfam; PF00037; fer4_2.
CC Pfam; PF00143; Radical_activat; 1.
CC ProDom; PD004758; Radical_activat; 1.
CC PROSITE; PS00198; 4FE4S-FERREDOXIN; 2.
CC PROSITE; PS01087; RADICAL_ACTIVATING; 1.
CC Hypothetical protein; Iron-sulfur; 4fe-4s; Complete proteome.
FT METAL 31 31 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 35 35 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 38 38 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT METAL 47 47 IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).

FT TRANSMEM 847 872 POTENTIAL.
FT TRANSMEM 914 934 POTENTIAL.
FT TRANSMEM 951 976 POTENTIAL.
FT MOD_RES 374 374 PHOSPHORYLATION (PROBABLE).
FT BINDING 506 506 ATP (BY SIMILARITY).
FT CONFLICT 109 109 C -> Y (IN REF. 2).
FT CONFLICT 185 185 I -> V (IN REF. 2).
FT CONFLICT 249 249 K -> E (IN REF. 2).
FT CONFLICT 309 309 L -> F (IN REF. 2).
SQ SEQUENCE 1021 AA; 112666 MW; 938A19AA487CBEAA CRC64;

Query Match 39.1%; Score 43; DB 1; Length 1021;
Best Local Similarity 47.1%; Pred No. 36;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 TVGFLKQDWGATIKD 20
   : ||: ||| | |
Db 879 THLLGLRVMDRWDIND 895

RESULT 12
YMI6_MYCLE
ID YMI6_MYCLE STANDARD; PRT; 307 AA.
AC O32960;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 32.9 KDA PROTEIN ML0860.
GN ML0860 OR MLCB22.18.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OC NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN.
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Felwell I., Fraser A., Hamlin N.,
RA Holroyd S., Horsby F., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.-A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RA "Massive gene decay in the leprosy bacillus."
RT Nature 409:1007-1011(2001).
RL Nature 409:1007-1011(2001).
CC -1- SIMILARITY: BELONGS TO THE UPF0105 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z98741; CAB11383.1; .
CC EMBL; AL583920; CAC31241.1; .
CC Leprona; ML0860; .
CC KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 307 AA; 32853 MW; EFFE5BA102455DA2 CRC64;

Query Match 38.2%; Score 42; DB 1; Length 307;
Best Local Similarity 70.0%; Pred No. 14;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 9 LKQDWGATGI 18
   : |||: |||:
Db 143 LQDWEGATL 152

```



```
FT PROPEP 1 5 BY SIMILARITY.
FT CHAIN 6 1020 SODIUM/POTASSIUM-TRANSPORTING ATPASE
FT ALPHA-2 CHAIN.
FT TRANSMEM 94 115 POTENTIAL.
FT TRANSMEM 128 147 POTENTIAL.
FT TRANSMEM 289 311 POTENTIAL.
FT TRANSMEM 318 346 POTENTIAL.
FT TRANSMEM 784 807 POTENTIAL.
FT TRANSMEM 846 871 POTENTIAL.
FT TRANSMEM 913 933 POTENTIAL.
FT TRANSMEM 950 975 POTENTIAL.
FT MOD_RES 374 374 PHOSPHORYLATION (PROBABLE).
FT BINDING 505 505 ATP (BY SIMILARITY).
SQ SEQUENCE 1020 AA; 112265 MW; AFBD8EA94FFB4FC3 CRC64;

Query Match 38.2%; Score 42; DB 1; Length 1020;
Best Local Similarity 40.0%; Pred. No. 53;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 6 VFLGKQDWGATIKD 20
   : | : | | | | : |
Db 880 LLGRLDWDORTMND 894

RESULT 15
GCST_NEIMB
ID GCST_NEIMB STANDARD; PRT; 366 AA.
AC Q9K0L8;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE AMINOMETHYLTRANSFERASE (EC 2.1.2.10) (GLYCINE CLEAVAGE SYSTEM
DE T. PROTEIN).
DE GCVT OR NMB0574.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175735; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
RC -|- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF
CC GLYCINE (BY SIMILARITY).
CC -|- CATALYTIC ACTIVITY: (6S)-TETRAHYDROFOLATE +
CC S-AMINOMETHYLDIHYDROLIPOYLPROTEIN = (6R)-5,10-
CC METHYLENETETRAHYDROFOLATE + NH(3) + DIHYDROLIPOYLPROTEIN
CC -|- SUBUNIT: THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF FOUR PROTEINS:
CC P. T. L. AND H (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE GCVT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE002413; AAF41002.1;
CC TIGR; NMB0574;
CC InterPro; IPR002536; GCV_T.
```

```
DR Pfam; PF01571; GCV_T; 1.
KW Transferase; Aminotransferase; Complete proteome.
SQ SEQUENCE 366 AA; 39740 MW; 566A7BCD21691D91 CRC64;

Query Match 37.7%; Score 41.5; DB 1; Length 366;
Best Local Similarity 44.8%; Pred. No. 21;
Matches 13; Conservative 2; Mismatches 3; Indels 11; Gaps 2;

QY 2 TTTTGVF--GLKQ-----DWDGATIK 19
   ||:||||| ||| |::||| |
Db 312 TTSGVFSPLKQSIARVPKDFDGTAK 340

Search completed: March 14, 2002, 09:24:49
Job time: 905 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model
Run on: March 14, 2002, 09:10:41 ; Search time 56.96 Seconds
(without alignments)
26.747 Million cell updates/sec

Title: US-09-765-739A-2
Perfect score: 110
Sequence: 1 NTTTGVFLKQDWDGATIKD 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	81.8	280	2 JE0217	28k surface antige
2	82	74.5	276	2 JE0218	28k surface antige
3	71	64.5	286	2 JE0219	28k surface antige
4	59	53.6	278	2 JE0216	28k surface antige
5	59	53.6	284	2 I40882	major antigenic pr
6	52	47.3	5188	2 B85477	probable RTX famil
7	51	46.4	1191	2 A53491	bumetanide-sensiti
8	49	44.5	133	2 JE0221	28k surface antige
9	47	42.7	540	2 S54586	probable membrane
10	47	42.7	584	2 C48658	flagellin - Escher
11	46.5	42.3	267	2 C83242	conserved hypotet
12	46	41.8	261	2 G84057	hypothetical prote
13	45	40.9	160	2 A75466	2-demethylmenaquin
14	45	40.9	756	2 T20109	hypothetical prote
15	44	40.0	284	2 B6945	phage-related prot
16	44	40.0	465	1 S47738	cytochrome-c perox
17	44	40.0	465	2 G86034	cytochrome-c perox
18	44	40.0	534	2 C82096	aminoacyl-histidin
19	44	40.0	648	1 P38PF6	p3 protein - phage
20	44	40.0	1004	2 JH0470	Na+/K+-exchanging
21	43.5	39.5	290	2 S76787	hypothetical prote
22	43	39.1	280	2 D70976	hypothetical prote
23	43	39.1	427	2 I40167	dihydroorotase (EC
24	43	39.1	470	2 JC4098	tetracycline 6-hyd
25	43	39.1	482	2 G75483	probable leucyl am
26	43	39.1	1037	2 A56594	Na+/K+-exchanging
27	43	39.1	1649	2 C86822	hypothetical prote
28	42	38.2	224	2 C72390	hypothetical prote
29	42	38.2	287	1 S56603	probable pyruvate

30 42 38.2 287 2 D86137 probable activation
31 42 38.2 307 2 T44893 hypothetical prote
32 42 38.2 362 2 F75379 S-adenosylmethioni
33 42 38.2 413 2 T08297 conserved hypotet
34 42 38.2 656 2 A41870 dnaA protein - Str
35 42 38.2 725 2 T00492 hypothetical prote
36 42 38.2 1020 2 A34474 Na+/K+-exchanging
37 41.5 37.7 176 2 B64549 conserved hypotet
38 41.5 37.7 293 2 A84110 sugar ABC transpor
39 41.5 37.7 366 2 A81183 glycine cleavage s
40 41.5 37.7 368 2 G81919 probable aminometh
41 41.5 37.7 468 2 T23091 hypothetical prote
42 41 37.3 95 2 B85637 hypothetical prote
43 41 37.3 244 2 B84885 hypothetical prote
44 41 37.3 269 2 A69997 hypothetical prote
45 41 37.3 376 2 A75381 peptide ABC transp

ALIGNMENTS

RESULT 1

JE0217
28k surface antigen 4 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0217
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0217
A:Molecule type: DNA
A:Residues: 1-280 <RED>
A:Cross-references: GB:AF062761

Query Match 81.8%; Score 90; DB 2; Length 280;
Best Local Similarity 88.9%; Pred. No. 1.3e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NTTTGVFLKQDWDGATI 18
Db 60 NTTTGVFLKQDWDGSTI 77
||| ||||| ||||| |||

RESULT 2

JE0218
28k surface antigen 5 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0218
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0218
A:Molecule type: DNA
A:Residues: 1-276 <RED>
A:Cross-references: GB:AF062761

Query Match 74.5%; Score 82; DB 2; Length 276;
Best Local Similarity 70.0%; Pred. No. 2.3e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NTTTGVFLKQDWDGATIKD 20
Db 59 NTTTGVFLKQDWDGSAISN 78
||| ||||| ||||| |||

```

RESULT 3
JE0219
28k surface antigen 2 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0219
R:Reddy, G.R.; Sulisana, C.R.; Barbet, A.F.; Mahan, S.M.; Burr ridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0219
A:Molecule type: DNA
A:Residues: 1-286 <RED>
A:Cross-references: GB:AF062761

Query Match 64.5%; Score 71; DB 2; Length 286;
Best Local Similarity 66.7%; Pred. No. 0.0013;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGAT 18
Db 60 NTTGVFGLKQDWGAT 77

RESULT 4
JE0216
28k surface antigen 3 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0216
R:Reddy, G.R.; Sulisana, C.R.; Barbet, A.F.; Mahan, S.M.; Burr ridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0216
A:Molecule type: DNA
A:Residues: 1-278 <RED>
A:Cross-references: GB:AF062761

Query Match 53.6%; Score 59; DB 2; Length 278;
Best Local Similarity 60.0%; Pred. No. 0.094;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWG 15
Db 60 NPTVALYGLKDWES 74

RESULT 5
I40882
major antigenic protein - heartwater rickettsia
C:Species: Cowdria ruminantium (heartwater rickettsia)
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40882; S42827
R:van Vliet, A.H.; Jongejan, F.; van Kleef, M.; van der Zeijst, B.A.
Infect. Immun. 62, 1451-1456, 1994
A:Title: Molecular cloning, sequence analysis, and expression of the gene encoding the
A:Reference number: I40882; MUID:94178956
A:Accession: I40882
A:Molecule type: DNA
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-284 <RES>
A:Cross-references: EMBL:X74250; NID:g454266; PIDN:CAA52309.1; PID:g454267
C:Genetics:
A:Gene: map1

Query Match 53.6%; Score 59; DB 2; Length 284;
Best Local Similarity 76.9%; Pred. No. 0.096;

```

```

Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTGVFGLKQDWG 15
Db 63 TKAVFGLKQDWG 75

RESULT 6
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B85547
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: B85480; MUID:21074935; PMID:11206551
A:Accession: B85547
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0615

Query Match 47.3%; Score 52; DB 2; Length 5188;
Best Local Similarity 60.0%; Pred. No. 28;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 TTGVFGLKQDWGA 16
Db 4803 TTGVAAMDYWDGA 4817

RESULT 7
A53491
bumetanide-sensitive Na-K-Cl cotransporter - spiny dogfish
C:Species: Squalus acanthias (spiny dogfish)
C:Date: 02-Jun-1994 #sequence_revision 02-Jun-1994 #text_change 02-Mar-2001
C:Accession: A53491
R:Xu, J.C.; Lytle, C.; Zhu, T.T.; Payne, J.A.; Benz Jr., E.; Forbush III, B.
Proc. Natl. Acad. Sci. U.S.A. 91, 2201-2205, 1994
A:Title: Molecular cloning and functional expression of the bumetanide-sensitive Na-K
A:Reference number: A53491; MUID:94181560
A:Accession: A53491
A:Molecule type: mRNA
A>Status: preliminary
A:Residues: 1-1191 <XUA>
A:Cross-references: GB:U05958; NID:g454096; PIDN:AAB60617.1; PID:g454097
C:Superfamily: rat bumetanide-sensitive Na+/K+/Cl--cotransport protein

Query Match 46.4%; Score 51; DB 2; Length 1191;
Best Local Similarity 58.8%; Pred. No. 8.2;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 TGVFGLKQDWGATIKD 20
Db 856 TLVFGKKDWQALMKD 872

RESULT 8
JE0221
28k surface antigen 2 - Ehrlichia canis
C:Species: Ehrlichia canis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 21-Jul-2000
C:Accession: JE0221
R:Reddy, G.R.; Sulisana, C.R.; Barbet, A.F.; Mahan, S.M.; Burr ridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri

```

A:Reference number: JE0216; MUID:98321180
A:Accession: JE0221
A:Molecule type: DNA
A:Residues: 1-133 <RED>
A:Cross-references: GB:AF062762; NID:g3327964; PIDN:AAC26722.1; PID:g3327966

Query Match 44.5%; Score 49; DB 2; Length 133;
Best Local Similarity 56.2%; Pred. No. 1.6;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 TTGVFGLKQDWDGATI 18
|||:||||:| | |
DB 62 TTVVYGLKNWAGDAI 77

RESULT 9
S54586

probable membrane protein YMR279c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YMR021.05c

C:Species: Saccharomyces cerevisiae

C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 05-Nov-1999

C:Accession: S54586

R:Pearson, D.; Bowman, S.

submitted to the EMBL Data Library, May 1995

A:Reference number: S54582

A:Accession: S54586

A:Molecule type: DNA

A:Residues: 1-540 <PEA>

A:Cross-references: EMBL:Z49704; NID:g825540; PIDN:CAA89777.1; PID:g825545; GSPDB:GN0001

A:Experimental source: strain AB972

C:Genetics:

A:Gene: MIPS:YMR279c

A:Map position: 13R

C:Superfamily: aminotriazole resistance protein YML116W

C:Keywords: transmembrane protein

F:109-125/Domain: transmembrane #status predicted <TM1>

F:139-155/Domain: transmembrane #status predicted <TM2>

F:174-190/Domain: transmembrane #status predicted <TM3>

F:233-249/Domain: transmembrane #status predicted <TM4>

F:268-284/Domain: transmembrane #status predicted <TM5>

F:298-314/Domain: transmembrane #status predicted <TM6>

F:333-351/Domain: transmembrane #status predicted <TM7>

F:377-393/Domain: transmembrane #status predicted <TM8>

F:400-416/Domain: transmembrane #status predicted <TM9>

F:435-451/Domain: transmembrane #status predicted <TM10>

F:503-519/Domain: transmembrane #status predicted <TM11>

Query Match 42.7%; Score 47; DB 2; Length 540;
Best Local Similarity 38.9%; Pred. No. 15;
Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGATI 18
| | : || | | : |
DB 256 NVPTNIHGLSMDWTGSAL 273

RESULT 10
C48658

flagellin - Escherichia coli (strain U5-41)

C:Species: Escherichia coli

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 26-Aug-1999

C:Accession: C48658

R:Schoenhals, G.; Whitfield, C.

J. Bacteriol. 175, 5395-5402, 1993

A:Title: Comparative analysis of flagellin sequences from Escherichia coli strains posse

A:Reference number: A48658; MUID:93374833

A:Accession: C48658

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-584 <SCH>

A:Cross-references: GB:L07388; NID:g290438; PIDN:AAA23798.1; PID:g290439

C:Superfamily: flagellin

Query Match 42.7%; Score 47; DB 2; Length 584;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQD 12
|||||:||||:
DB 199 NTTGLYGLKTE 210

RESULT 11
C83242

conserved hypothetical protein PA3239 [imported] - Pseudomonas aeruginosa (strain PAO

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: C83242

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: C83242

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-267 <STO>

A:Cross-references: GB:AE004746; GB:AE004091; NID:g9949350; PIDN:AAG06627.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA3239

Query Match 42.3%; Score 46.5; DB 2; Length 267;
Best Local Similarity 64.7%; Pred. No. 8.2;
Matches 11; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

QY 1 NTTGVFGLKQDWDGAT 17
|||||:| | | | | |
DB 145 NTTFGVLGL--WDPAT 158

RESULT 12

G84057

hypothetical protein BH3263 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000

C:Accession: G84057

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20263314

A:Accession: G84057

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-261 <STO>

A:Cross-references: GB:AP001518; GB:BA000004; NID:g10175792; PIDN:BA006982.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3263

C:Superfamily: Bacillus subtilis hypothetical protein ytmP

Query Match 41.8%; Score 46; DB 2; Length 261;
Best Local Similarity 77.8%; Pred. No. 9.5;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 12 DWDGATIKD 20
|||||: |
DB 177 DWDGATVAD 185

RESULT 13

A75466
 2-demethylmenaquinone 2-C-methyltransferase (EC 2.1.1.-) DR0859 [similarity] - Deinococcus
 C:Species: Deinococcus radiodurans
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
 C:Accession: A75466
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: A75466
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-160 <WHI>
 A:Cross-references: GB:AE001940; GB:AE000513; NID:g6458577; PIDN:AAF10437.1; PID:g645857
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0859
 A:Map position: 1
 C:Keywords: methyltransferase

Query Match 40.9%; Score 45; DB 2; Length 160;
 Best Local Similarity 50.0%; Pred. No. 8;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 GVFGKQDWGATG 18
 ||||: |||
 Db 80 GVFGVNGWGVII 93

RESULT 14

T20109
 Hypothetical protein C50F4.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
 C:Accession: T20109
 R:McMurray, A.
 submitted to the EMBL Data Library, April 1996
 A:Reference number: Z19225
 A:Accession: T20109
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-756 <WIL>
 A:Cross-references: EMBL:Z70750; PIDN:CAA94737.1; GSPDB:GN00023; CESP:C50F4.2
 A:Experimental source: clone C50F4
 C:Genetics:
 A:Gene: C50F4.2
 A:Map position: 5
 A:Introns: 24/1; 245/3; 315/2; 389/3; 537/3; 568/2; 747/1
 C:Superfamily: human 6-phosphofructokinase; 6-phosphofructokinase 1 homology

Query Match 40.9%; Score 45; DB 2; Length 756;
 Best Local Similarity 53.3%; Pred. No. 43;
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 6 VFGLKQDWGATK 20
 ||||| ||
 Db 436 VIGIKHGWGLKNKD 450

RESULT 15

B69945
 phage-related protein homolog yqak - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 05-Dec-1999 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C:Accession: B69945
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berton
 C.; Brode, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parco, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, J.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Se
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Zumbstein, E.; Yata, K.; Yoshida
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
 A:Reference number: A69580; MUID:98044033
 A:Accession: B69945
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-284 <KUN>
 A:Cross-references: GB:Z99117; GB:AL009126; NID:g2634966; PIDN:CAB14569.1; PID:g26350
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yqak
 C:Superfamily: Escherichia coli recf protein

Query Match 40.0%; Score 44; DB 2; Length 284;
 Best Local Similarity 53.8%; Pred. No. 22;
 Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 7 FGKQDWGATK 19
 ||||| ||
 Db 193 FGKNDWDAMALK 205

Search completed: March 14, 2002, 09:10:42
 Job time: 358 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:07:39 ; Search time 54.54 Seconds
(without alignments)
8.252 Million cell updates/sec

Title: US-09-765-739A-3
Perfect score: 107
Sequence: 1 NTTGVFGLKQNWGSAISN 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	276	4	US-08-953-326-18
2	107	100.0	280	3	US-08-733-230-4
3	107	100.0	280	4	US-08-953-326-4
4	91	85.0	280	4	US-08-953-326-17
5	73	68.2	286	4	US-08-953-326-15
6	61.5	57.5	133	4	US-08-953-326-20
7	60	56.1	278	4	US-08-953-326-16
8	51	47.7	287	3	US-08-733-230-2
9	51	47.7	287	4	US-08-953-326-2
10	41	38.3	365	3	US-08-840-146-19
11	41	38.3	365	3	US-09-360-220-19
12	41	38.3	400	2	US-08-436-771-11
13	41	38.3	400	2	US-08-434-998-11
14	41	38.3	400	2	US-08-487-797-11
15	41	38.3	400	5	PCT-US95-02058-11
16	39	36.4	137	2	US-08-630-822A-64
17	39	36.4	137	2	US-09-005-069-64
18	39	36.4	388	2	US-08-742-621-1
19	39	36.4	388	4	US-09-191-608-22
20	39	36.4	527	1	US-08-462-484-10
21	39	36.4	527	1	US-08-441-147-10
22	39	36.4	527	5	PCT-US95-07536-10
23	39	36.4	2476	2	US-08-276-967-2
24	38	35.5	377	1	US-07-772-087-2
25	38	35.5	738	4	US-08-989-385-1
26	37	34.6	31	1	US-08-190-802A-70
27	37	34.6	31	1	US-08-190-802A-129

28	37	34.6	31	1	US-08-190-802A-173
29	37	34.6	31	4	US-08-477-346-70
30	37	34.6	31	4	US-08-477-346-129
31	37	34.6	31	4	US-08-477-346-173
32	37	34.6	317	1	US-08-190-802A-27
33	37	34.6	317	1	US-08-190-802A-41
34	37	34.6	317	1	US-08-190-802A-47
35	37	34.6	317	4	US-08-477-346-27
36	37	34.6	317	4	US-08-477-346-41
37	37	34.6	317	4	US-08-477-346-47
38	37	34.6	523	4	US-09-550-338-2
39	37	34.6	592	2	US-08-366-490-2
40	37	34.6	592	2	US-08-366-490-6
41	37	34.6	592	3	US-08-860-483A-2
42	37	34.6	592	3	US-08-860-483A-6
43	37	34.6	596	3	US-08-860-483A-12
44	37	34.6	596	3	US-08-860-483A-13
45	37	34.6	600	2	US-08-679-405-2

ALIGNMENTS

RESULT 1

US-08-953-326-18
; Sequence 18, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-18

Query Match 100.0%; Score 107; DB 4; Length 276;
Best Local Similarity 100.0%; Pred. No. 4.9e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQNWGSAISN 20
Db 59 NTTGVFGLKQNWGSAISN 78

RESULT 2

US-08-733-230-4
; Sequence 4, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.

APPLICANT: Mahan, Suman M.
TITLE OF INVENTION: Nucleic Acid Vaccines Against
Rickettsial Diseases and Methods of Use
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/733,230
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF-167
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 280 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-733-230-4

Query Match 100.0%; Score 107; DB 3; Length 280;
Best Local Similarity 100.0%; Pred. No. 5e-10; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

QY 1 NTTGVFGLKQNDGSAISN 20
Db 60 NTTGVFGLKQNDGSAISN 79

RESULT 3
US-08-953-326-4
; Sequence 4, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4

Query Match 100.0%; Score 107; DB 4; Length 280;
Best Local Similarity 100.0%; Pred. No. 5e-10; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

QY 1 NTTGVFGLKQNDGSAISN 20
Db 60 NTTGVFGLKQNDGSAISN 79

RESULT 4
US-08-953-326-17
; Sequence 17, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-17

Query Match 85.0%; Score 91; DB 4; Length 280;
Best Local Similarity 84.2%; Pred. No. 2.1e-07; Indels 1; Gaps 0;
Matches 16; Conservative 2; Mismatches 1;

QY 1 NTTGVFGLKQNDGSAIS 19
Db 60 NTTGVFGLKQNDGSTIS 78

RESULT 5
US-08-953-326-15
; Sequence 15, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-15

Query Match 68.2%; Score 73; DB 4; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00019;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 6
US-08-953-326-20
; Sequence 20, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 20
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-08-953-326-20

Query Match 57.5%; Score 61.5; DB 4; Length 133;
Best Local Similarity 68.4%; Pred. No. 0.006;
Matches 13; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 2 TTVGVFGLKQNWGSAIS 20
Db 62 TTV-VYGLKENWAGDAISS 79

RESULT 7
US-08-953-326-16
; Sequence 16, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326

; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 16
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-16

Query Match 56.1%; Score 60; DB 4; Length 278;
Best Local Similarity 50.0%; Pred. No. 0.025;
Matches 10; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAIS 20
Db 60 NPTVALYGLKQDWEGISS 79

RESULT 8
US-08-733-230-2
; Sequence 2, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-2

Query Match 47.7%; Score 51; DB 3; Length 287;
Best Local Similarity 69.2%; Pred. No. 0.77;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVGVFGLKQNWGDG 15

RESULT 12
US-08-436-771-11
; Sequence 11, Application US/08436771
; Patent No. 5861300
; GENERAL INFORMATION:
; APPLICANT: Silverman, Robert H.
; TITLE OF INVENTION: Antiviral Transgenic Plants, Vectors,
; TITLE OF INVENTION: Cells and Methods
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
; ADDRESSEE: Russell
; STREET: 200 E. Broward Boulevard
; CITY: Fort Lauderdale
; STATE: Florida
; COUNTRY: USA
; ZIP: 33301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,771
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/198,973
; FILING DATE: 18-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Manso, Peter J.
; REGISTRATION NUMBER: 32,264
; REFERENCE/DOCKET NUMBER: CL11363-16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 305/527/2498
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-436-771-11

Query Match 38.3%; Score 41; DB 2; Length 400;
Best Local Similarity 60.0%; Pred. No. 49;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 QNWDGSAIN 20
DB 333 KNWDGSPVSS 342
RESULT 13
US-08-434-998-11
; Sequence 11, Application US/08434998
; Patent No. 5866781
; GENERAL INFORMATION:
; APPLICANT: Silverman, Robert H.
; TITLE OF INVENTION: Antiviral Transgenic Plants, Vectors,
; TITLE OF INVENTION: Cells and Methods
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
; ADDRESSEE: Russell
; STREET: 200 E. Broward Boulevard
; CITY: Fort Lauderdale

STATE: Florida
COUNTRY: USA
ZIP: 33301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,998
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/198,973
FILING DATE: 18-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Manso, Peter J.
REGISTRATION NUMBER: 32,264
REFERENCE/DOCKET NUMBER: CL11363-16
TELECOMMUNICATION INFORMATION:
TELEPHONE: 305/527/2498
TELEFAX: 305/764/4996
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-434-998-11

Query Match 38.3%; Score 41; DB 2; Length 400;
Best Local Similarity 60.0%; Pred. No. 49;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 QNWDGSAIN 20
DB 333 KNWDGSPVSS 342
RESULT 14
US-08-487-797-11
; Sequence 11, Application US/08487797
; Patent No. 5866787
; GENERAL INFORMATION:
; APPLICANT: Silverman, Robert H.
; APPLICANT: Sengupta, Dibyendu N.
; TITLE OF INVENTION: Transgenic Plants Co-Expressing A
; TITLE OF INVENTION: Functional Human 2-5A System
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Holland & Knight
; STREET: One E. Broward Boulevard, #1300
; CITY: Fort Lauderdale
; STATE: Florida
; COUNTRY: USA
; ZIP: 33301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,797
FILING DATE: 07-JUNE-1995
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Manso, Peter J.
REGISTRATION NUMBER: 32,264
REFERENCE/DOCKET NUMBER: CL11363-16(C)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 305/468-7811

TELEFAX: 305/463-2030
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-487-797-11

Search completed: March 14, 2002, 09:07:40
Job time: 841 sec

Query Match 38.3%; Score 41; DB 2; Length 400;
Best Local Similarity 60.0%; Pred. No. 49;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 QNWDGSAISN 20
Db 333 KNWDGSPVSS 342

RESULT 15
PCT-US95-02058-11
; Sequence 11, Application PC/TUS9502058
; GENERAL INFORMATION:
; APPLICANT: Silverman, Robert H.
; APPLICANT: Sengupta, Dibyendu N.
; TITLE OF INVENTION: Antiviral Transgenic Plants, Vectors,
; TITLE OF INVENTION: Cells and Methods
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ruden, Barnett, McClosky, Smith, Schuster &
; ADDRESSEE: Russell
; STREET: 200 E. Broward Boulevard
; CITY: Fort Lauderdale
; STATE: Florida
; COUNTRY: USA
; ZIP: 33301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02058
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/198,973
; FILING DATE: 18-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Manso, Peter J.
; REGISTRATION NUMBER: 32,264
; REFERENCE/DOCKET NUMBER: CL11363-16
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 305/527/2498
; TELEFAX: 305/764/4996
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-02058-11

Query Match 38.3%; Score 41; DB 5; Length 400;
Best Local Similarity 60.0%; Pred. No. 49;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 QNWDGSAISN 20
Db 333 KNWDGSPVSS 342

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:07:41 ; Search time 54.54 Seconds
(without alignments)
7.427 Million cell updates/sec

Title: us-09-765-739a-6

Perfect score: 97
Sequence: 1 NPTVALYGLKQDWEGISS 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/PT05_COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	100.0	278	4	US-08-953-326-16
2	62	63.9	280	4	US-08-953-326-17
3	58	59.8	276	4	US-08-953-326-18
4	58	59.8	280	3	US-08-733-230-4
5	58	59.8	280	4	US-08-953-326-4
6	52	53.6	287	3	US-08-733-230-2
7	52	53.6	287	4	US-08-953-326-2
8	51	52.6	286	4	US-08-953-326-15
9	48	49.5	133	4	US-08-953-326-20
10	42	43.3	915	4	US-09-346-237-2
11	42	43.3	928	1	US-08-474-140-11
12	42	43.3	928	1	US-08-477-630-11
13	42	43.3	928	1	US-08-472-293-11
14	42	43.3	928	1	US-08-474-545-11
15	42	43.3	928	2	US-08-478-341-11
16	42	43.3	928	3	US-08-996-733-11
17	39.5	40.7	99	2	US-08-860-577-3
18	39	40.2	263	1	US-07-901-707-7
19	39	40.2	263	1	US-07-988-430-7
20	39	40.2	263	1	US-08-425-336-7
21	39	40.2	263	1	US-08-488-113B-7
22	39	40.2	263	1	US-08-477-484B-7
23	39	40.2	263	2	US-08-646-360-7
24	39	40.2	263	4	US-08-839-785-7
25	39	40.2	263	4	US-09-136-389-7
26	39	40.2	263	5	PCT-US92-09487-7
27	39	40.2	609	1	US-07-798-776-2

28	39	40.2	609	3	US-08-251-288A-2	Sequence 2, Appli
29	39	40.2	609	3	US-09-298-819A-2	Sequence 2, Appli
30	38.5	39.7	683	3	US-08-947-965-2	Sequence 2, Appli
31	38	39.2	738	4	US-08-989-385-1	Sequence 1, Appli
32	38	39.2	862	4	US-09-346-237-1	Sequence 1, Appli
33	38	39.2	1479	3	US-08-840-062-4	Sequence 4, Appli
34	37.5	38.7	368	5	PCT-US93-11703-24	Sequence 24, Appli
35	37.5	38.7	393	1	US-08-499-568-15	Sequence 15, Appli
36	37.5	38.7	393	2	US-08-793-958-15	Sequence 15, Appli
37	37.5	38.7	393	2	US-08-956-998-2	Sequence 2, Appli
38	37.5	38.7	503	4	US-08-740-223A-18	Sequence 18, Appli
39	37.5	38.7	675	3	US-08-947-965-76	Sequence 76, Appli
40	37	38.1	553	1	US-07-683-957B-3	Sequence 3, Appli
41	37	38.1	575	1	US-07-683-957B-2	Sequence 2, Appli
42	37	38.1	595	2	US-08-677-049-11	Sequence 11, Appli
43	37	38.1	820	4	US-09-313-677-21	Sequence 21, Appli
44	37	38.1	926	4	US-09-313-677-2	Sequence 2, Appli
45	37	38.1	933	4	US-09-313-677-19	Sequence 19, Appli

ALIGNMENTS

RESULT 1
US-08-953-326-16
; Sequence 16, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-16

Query Match 100.0%; Score 97; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 3.6e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NPTVALYGLKQDWEGISS 18
Db 60 NPTVALYGLKQDWEGISS 77

RESULT 2
US-08-953-326-17
; Sequence 17, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.

```

; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-17

Query Match      63.9%; Score 62; DB 4; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.0042;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
   | | :|||||:|
Db 60 NTTGVFGLKQNDWG 74

RESULT 3
US-08-953-326-18
; Sequence 18, Application US/089533326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-18

Query Match      59.8%; Score 58; DB 4; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.02;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
   | | :|||||:|
Db 59 NTTGVFGLKQNDWG 73

RESULT 4
US-08-733-230-4
; Sequence 4, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.

```

```

; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-17

Query Match      59.8%; Score 58; DB 3; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.021;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
   | | :|||||:|
Db 60 NTTGVFGLKQNDWG 74

RESULT 5
US-08-953-326-4
; Sequence 4, Application US/089533326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0

```



```

; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4

Query Match      59.8%; Score 58; DB 4; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.021;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWE 15
Db 60 NPTVGVLKQKWDG 74

RESULT 6
US-08-733-230-2
; Sequence 2, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-2

Query Match      53.6%; Score 52; DB 3; Length 287;
Best Local Similarity 43.8%; Pred. No. 0.24;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

```

```

US-08-953-326-2
; Sequence 2, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
US-08-953-326-2

Query Match      53.6%; Score 52; DB 4; Length 287;
Best Local Similarity 43.8%; Pred. No. 0.24;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEIGISS 18
Db 63 TQTVFGLKKWDGDKVT 78

RESULT 8
US-08-953-326-15
; Sequence 15, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-15

Query Match      52.6%; Score 51; DB 4; Length 286;
Best Local Similarity 50.0%; Pred. No. 0.35;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 NPTVALYGLKQDWE 14

```

Matches	9;	Conservative	3;	Mismatches	4;	Indels	2;	Gaps	1;
QY	1	NPTVALYGLKQDWEGISS 18							
Db	628	NFGIALYG--EPWTGTS 643							
RESULT	11								
US-08-474-140-11									
Sequence 11, Application US/08474140									
Patent No. 5721127									
GENERAL INFORMATION:									
APPLICANT: DEWEER, PHILIPPE									
APPLICANT: AMORY, ANTOINE									
TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH									
TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS									
TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF									
NUMBER OF SEQUENCES: 15									
CORRESPONDENCE ADDRESS:									
ADDRESSEE: WILLIAN BRINKS HOFER GILSON & LIONE, P.C.									
STREET: 2000 K Street, N.W., Suite 200									
CITY: Washington									
STATE: D.C.									
COUNTRY: U.S.A.									
ZIP: 20006									
COMPUTER READABLE FORM:									
MEDIUM TYPE: Floppy disk									
COMPUTER: IBM PC compatible									
OPERATING SYSTEM: PC-DOS/MS-DOS									
SOFTWARE: PatentIn Release #1.0, Version #1.25									
CURRENT APPLICATION DATA:									
APPLICATION NUMBER: US/08/474,140									
FILING DATE: 07-JUN-1995									
CLASSIFICATION: 435									
ATTORNEY/AGENT INFORMATION:									
NAME: Wilhem F. Gadiano, Esq.									
REGISTRATION NUMBER: 37,136									
REFERENCE/DOCKET NUMBER: 4121-41									
TELECOMMUNICATION INFORMATION:									
TELEPHONE: (202) 429-0625									
TELEFAX: (202) 293-1850									
TELEX: 650 383-5605									
INFORMATION FOR SEQ ID NO: 11:									
SEQUENCE CHARACTERISTICS:									
LENGTH: 928 amino acids									
TYPE: amino acids									
TOPOLOGY: linear									
MOLECULE TYPE: protein									
US-08-474-140-11									
Query Match	43.3%	Score 42:	DB 1;	Length 928;					
Best Local Similarity	50.0%	Pred. No. 52:							
Matches	9;	Conservative	3;	Mismatches	4;	Indels	2;	Gaps	1;
QY	1	NPTVALYGLKQDWEGISS 18							
Db	641	NFGIALYG--EPWTGTS 656							
RESULT	12								
US-08-477-630-11									
Sequence 11, Application US/08477630									
Patent No. 5721128									
GENERAL INFORMATION:									
APPLICANT: DEWEER, PHILIPPE									
APPLICANT: AMORY, ANTOINE									
TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH									
TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS									
TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF									
NUMBER OF SEQUENCES: 15									
CORRESPONDENCE ADDRESS:									
ADDRESSEE: WILLIAN BRINKS HOFER GILSON & LIONE, P.C.									

STREET: 2000 K Street, N.W., Suite 200
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20006
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/477,630
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Wilhem F. Gadiano, Esq.
 REGISTRATION NUMBER: 37,136
 REFERENCE/DOCKET NUMBER: 4121-42
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 429-0625
 TELEFAX: (202) 293-1850
 TELEX: 650 383-5605
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 928 amino acids
 TYPE: amino acids
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-477-630-11

Query Match 43.3%; Score 42; DB 1; Length 928;
 Best Local Similarity 50.0%; Pred. No. 52;
 Matches 9; Conservative 3; Mismatches 4; Indels 2; Gaps 1;
 QY 1 NPTVALYGLKQDWEGISS 18
 Db 641 NFGIALYG--EPWTGGTS 656

RESULT 13
 US-08-472-293-11
 ; Sequence 11, Application US/08472293
 ; Patent No. 5731174
 ; GENERAL INFORMATION:
 ; APPLICANT: DEWEER, PHILIPPE
 ; TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH
 ; TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS
 ; TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: WILLIAM BRINKS HOFER GILSON & LIONE, P.C.
 ; STREET: 2000 K Street, N.W., Suite 200
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20006
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/472,293
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Wilhem F. Gadiano, Esq.
 ; REGISTRATION NUMBER: 37,136
 ; REFERENCE/DOCKET NUMBER: 4121-44
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 429-0625

TELEFAX: (202) 293-1850
 TELEX: 650 383-5605
 INFORMATION FOR SEQ ID NO: 11:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 928 amino acids
 TYPE: amino acids
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-472-293-11

Query Match 43.3%; Score 42; DB 1; Length 928;
 Best Local Similarity 50.0%; Pred. No. 52;
 Matches 9; Conservative 3; Mismatches 4; Indels 2; Gaps 1;
 QY 1 NPTVALYGLKQDWEGISS 18
 Db 641 NFGIALYG--EPWTGGTS 656

RESULT 14
 US-08-474-545-11
 ; Sequence 11, Application US/08474545
 ; Patent No. 5736375
 ; GENERAL INFORMATION:
 ; APPLICANT: DEWEER, PHILIPPE
 ; APPLICANT: AMORY, ANTOINE
 ; TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH
 ; TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS
 ; TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: WILLIAM BRINKS HOFER GILSON & LIONE, P.C.
 ; STREET: 2000 K Street, N.W., Suite 200
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20006
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/474,545
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Wilhem F. Gadiano, Esq.
 ; REGISTRATION NUMBER: 37,136
 ; REFERENCE/DOCKET NUMBER: 4121-43
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 429-0625
 ; TELEFAX: (202) 293-1850
 ; TELEX: 650 383-5605
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 928 amino acids
 ; TYPE: amino acids
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-474-545-11

Query Match 43.3%; Score 42; DB 1; Length 928;
 Best Local Similarity 50.0%; Pred. No. 52;
 Matches 9; Conservative 3; Mismatches 4; Indels 2; Gaps 1;
 QY 1 NPTVALYGLKQDWEGISS 18
 Db 641 NFGIALYG--EPWTGGTS 656

```

RESULT 15
US-08-478-341-11
; Sequence 11, Application US/08478341
; Patent No. 5817498
; GENERAL INFORMATION:
; APPLICANT: DEWEER, PHILIPPE
; APPLICANT: AMORY, ANTOINE
; TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH
; TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS
; TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WILLIAM BRINKS HOFER GILSON & LIONE, P.C.
; STREET: 2000 K Street, N.W., Suite 200
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,341
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilhem F. Gadiano, Esq.
; REGISTRATION NUMBER: 37,136
; REFERENCE/DOCKET NUMBER: 4121-45
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 429-0625
; TELEFAX: (202) 293-1850
; TELEX: 650 383-5605
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 928 amino acids
; TYPE: amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-478-341-11

```

```

Query Match 43.3%; Score 42; DB 2; Length 928;
Best Local Similarity 50.0%; Pred. No. 52;
Matches 9; Conservative 3; Mismatches 4; Indels 2; Gaps 1;

QY 1 NPTVALYGLKQDWEGISS 18
   ||:||||:|:|:|
Db 641 NPGIALYG--EPWTGTS 656

Search completed: March 14, 2002, 09:07:42
Job time: 843 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:09:38 ; Search time 111.55 Seconds
(without alignments)
11.953 Million cell updates/sec

Title: US-09-765-739A-6

Perfect score: 97
Sequence: 1 NPTVALYGLKQDWEGISS 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97	100.0	278	19 AAW51093	Ehrlichia chaffeen
2	97	100.0	278	20 AAY06947	E. chafeensis OMP-
3	97	100.0	278	21 AAB36187	Ehrlichia chaffeen
4	97	100.0	278	22 AAU04197	Variable surface a
5	88	90.7	280	20 AAY06945	E. chafeensis OMP-
6	62	63.9	280	19 AAW51094	Ehrlichia chaffeen
7	62	63.9	280	21 AAB36188	Ehrlichia chaffeen
8	62	63.9	280	22 AAU04198	Variable surface a
9	59	60.8	280	20 AAY06948	E. chafeensis OMP-
10	59	60.8	288	20 AAY06959	E. canis P30 prote
11	58	59.8	256	20 AAY06942	E. chafeensis p28

12	58	59.8	276	19 AAW51095	Ehrlichia chaffeen
13	58	59.8	276	21 AAB36189	Ehrlichia chaffeen
14	58	59.8	276	22 AAU04199	Variable surface a
15	58	59.8	280	19 AAW51089	Ehrlichia chaffeen
16	58	59.8	280	21 AAB36183	Ehrlichia chaffeen
17	58	59.8	280	22 AAU04193	Major antigenic pr
18	58	59.8	281	20 AAY06943	E. chafeensis OMP-
19	56	57.7	280	20 AAY06962	E. canis P30-2 pro
20	56	57.7	280	21 AAY71479	Ehrlichia canis im
21	53	54.6	278	21 AAY71477	Ehrlichia canis im
22	53	54.6	307	20 AAY06961	E. canis P30-1 pro
23	52	53.6	287	19 AAW51088	Cowdria ruminantiu
24	52	53.6	287	21 AAB36182	Major antigenic pr
25	52	53.6	287	22 AAU04192	Ehrlichia chaffeen
26	51	52.6	286	19 AAW51092	E. chafeensis OMP-
27	51	52.6	286	20 AAY06946	Ehrlichia chaffeen
28	51	52.6	286	21 AAB36186	Variable surface a
29	51	52.6	286	22 AAU04196	Variable surface a
30	48	49.5	132	22 AAU04201	Ehrlichia canis VS
31	48	49.5	133	19 AAW51097	Ehrlichia canis pa
32	48	49.5	133	21 AAB36191	Ehrlichia canis im
33	48	49.5	133	21 AAY71480	E. canis P30-4 pro
34	48	49.5	276	20 AAY06964	Ehrlichia canis im
35	48	49.5	283	21 AAY71478	Ehrlichia canis im
36	47	48.5	904	20 AAY21976	Senesence-associat
37	43	44.3	50	21 AAB33357	Pinus radiata tran
38	43	44.3	53	21 AAB33358	Pinus radiata tran
39	43	44.3	237	21 AAB33257	Pinus radiata tran
40	42	43.3	187	21 AAY81775	Streptococcus pneu
41	42	43.3	815	22 AAE05695	Bacillus deramific
42	42	43.3	817	22 AAE05694	Bacillus deramific
43	42	43.3	826	20 AAY31748	Bacillus deramific
44	42	43.3	830	20 AAY31747	Bacillus deramific
45	42	43.3	915	21 AAY78514	Pullulanase amino

ALIGNMENTS

RESULT 1		AAW51093 standard; Protein; 278 AA.	
ID	AAW51093	standard; Protein; 278 AA.	
AC	AAW51093;		
DT	14-SEP-1998	(first entry)	
DE	Ehrlichia chafeensis VSA3 protein.		
KW	MAP1 homologue; variable surface antigen; VSA3; rickettsia;		
KW	DNA vaccine.		
OS	Ehrlichia chafeensis.		
XX			
EH	Key	Location/Qualifiers	
FT	Peptide	1-25	
FT		/note="putative signal peptide"	
XX			
PN	WO9816554-A1.		
PD	23-APR-1998.		
XX			
PF	17-OCT-1997;	97WO-US19044.	
XX			
PR	17-OCT-1996;	96US-0733230.	
XX			
PA	(UYFL) UNIV FLORIDA.		
XX			
PI	Barbet-AF, Burrledge MJ, Ganta RR, McGuire TC;		
PI	Nyika A, Rurangirwa FR;		
XX			
DR	WPI; 1998-251232/22.		
DR	N-PSDB; AAV07179.		

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
|||||
Db 60 nptvalyglkqdwegiss 77

RESULT 4
AAU04197
ID AAU04197 standard; Protein; 278 AA.

AC AAU04197;

DT 23-OCT-2001 (first entry)

DE Variable surface antigen 3 (VSA3) from Ehrlichia chaffeensis.

KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
KW infection; heartwater; diagnostic; variable surface antigen; VSA.

OS Ehrlichia chaffeensis.

PN US6251872-B1.

XX 26-JUN-2001.

PF 17-OCT-1997; 97US-0953326.

PR 17-OCT-1996; 96US-0733230.

XX (UYFL) UNIV FLORIDA.

PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;

PI Rutangirwa FR, Mahan SM, Bowie MV, Alleman AR;

DR WPI; 2001-424487/45.

DR N-PSDB; AAS07578.

PT New MAP2 genes and polypeptides useful as vaccines for conferring
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
PT as molecular markers in nucleic acid analysis procedures

XX Example 3; Fig 2A-2B; 30pp; English.

PS The sequence represents the amino acid sequence of variable surface
CC antigen 3 (VSA3) isolated from Ehrlichia chaffeensis, which
CC has similarity to major antigen protein (MAP). The MAP polynucleotides
CC and polypeptides are useful as vaccines for conferring immunity to
CC rickettsia infection, including Cowdria ruminantium causing heartwater.
CC The MAP polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures, and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish infectious
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
CC needed.

XX Sequence 278 AA;

Query Match 100.0%; Score 97; DB 22; Length 278;
Best Local Similarity 100.0%; Pred. No. 2.2e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
|||||
Db 60 nptvalyglkqdwegiss 77

RESULT 5
AAU06945

ID AAU06945 standard; Protein; 280 AA.

AC AAU06945;

DT 05-JUL-1999 (first entry)

DE E. chaffeensis OMP-1C protein.

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.

XX Ehrlichia chaffeensis.

PN WO9913720-A1.

XX 25-MAR-1999.

PF 18-SEP-1998; 98WO-US19600.

PR 19-SEP-1997; 97US-0059353.

PA (OHIS) UNIV OHIO STATE.

XX Ohashi N, Rikihisa Y;

PI WPI; 1999-254290/21.

DR N-PSDB; AAX34745.

PT Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis

PS Claim 13; Fig 5B; 55pp; English.

CC The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAU06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAU06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 280 AA;

Query Match 90.7%; Score 88; DB 20; Length 280;
Best Local Similarity 83.3%; Pred. No. 7.3e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
|||||
Db 60 nptvalyglkqdwgvs 77

RESULT 6

AAW51094

ID AAW51094 standard; Protein; 280 AA.

AC AAW51094;

DT 14-SEP-1998 (first entry)

DE Ehrlichia chaffeensis VSA4 protein.

KW MAP1 homologue; variable surface antigen; VSA4; rickettsia;
KW DNA vaccine.

XX Ehrlichia chaffeensis.

FF Key Location/Qualifiers
FT Peptide 1..25
FT /note= "putative signal peptide"

XX WO9816554-A1.

XX

PD 23-APR-1998.
 XX
 PF 17-OCT-1997; 97WO-US19044.
 XX
 PR 17-OCT-1996; 96US-0733230.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 XX
 DR WPI: 1998-251232/22.
 DR N-PSDB; AAV07179.
 XX
 PT Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX
 PS Claim 3; Fig 2B; 39pp: English.
 XX
 CC This is the full-length variable surface antigen VSA4 protein of
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
 CC partial open reading frame (ORF4) of a genomic locus (see AAV07179)
 CC of E. chaffeensis that was obtained on the basis of homology to the
 CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
 CC This genomic locus included 5 ORFs encoding similar, but
 CC non-identical proteins (see AAW51091-95). A claimed composition
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
 CC (see AAW51088-99) that elicits a protective immune response against a
 CC rickettsial pathogen. The nucleic acid is used, in human or
 CC veterinary medicine, in vaccines to protect against Rickettsia,
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
 CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).
 XX
 SQ Sequence 280 AA;

Query Match 63.9%; Score 62; DB 19; Length 280;
 Best Local Similarity 60.0%; Pred. NO. 0.018; 2; Indels 0; Gaps 0;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEQ 15
 | | : : : : : : : :
 Db 60 ntigvfglkqdwgdg 74

RESULT 7
 AAB36188
 ID AAB36188 standard; Protein; 280 AA.
 XX
 AC AAB36188;
 XX
 DT 02-MAR-2001 (first entry)
 XX
 DE Ehrlichia chaffeensis partial VSA4.
 XX
 KW Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAP1;
 KW major antigenic protein 1; antirickettsial vaccine; gene therapy;
 KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
 KW 3gdorf3.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN WO200065063-A2.
 XX
 PD 02-NOV-2000.
 XX
 PF 21-APR-2000; 2000WO-US10886.
 XX
 PR 22-APR-1999; 99US-0130725.
 XX
 PA (UYFL) UNIV FLORIDA.

XX
 PI Barbet AF, Bowle MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
 XX
 DR WPI: 2000-679675/66.
 DR N-PSDB; AAC68705.
 XX
 PT New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
 XX
 PS Claim 3; Page 45-46; 63pp: English.
 XX
 CC The present sequence shows a high degree of similarity to the major
 CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
 CC used in a vaccines to protect animals or humans against rickettsial
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid vaccines
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
 CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.
 XX
 SQ Sequence 280 AA;

Query Match 63.9%; Score 62; DB 21; Length 280;
 Best Local Similarity 60.0%; Pred. NO. 0.018; 2; Indels 0; Gaps 0;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEQ 15
 | | : : : : : : : :
 Db 60 ntigvfglkqdwgdg 74

RESULT 8
 AAU04198
 ID AAU04198 standard; Protein; 280 AA.
 XX
 AC AAU04198;
 XX
 DT 23-OCT-2001 (first entry)
 XX
 DE Variable surface antigen 4 (VSA4) from Ehrlichia chaffeensis.
 XX
 KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
 KW infection; heartwater; diagnostic; variable surface antigen; VSA.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN US6251872-B1.
 XX
 PD 26-JUN-2001.
 XX
 PF 17-OCT-1997; 97US-0953326.
 XX
 PR 17-OCT-1996; 96US-0733230.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
 PI Rurangirwa FR, Mahan SM, Bowle MV, Alleman AR;
 XX
 DR WPI: 2001-424487/45.
 DR N-PSDB; AAS07578.
 XX
 PT New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures

XX PS Example 3; Fig 2A-2B; 30pp; English.

XX CC The sequence represents the amino acid sequence of variable surface

CC antigen 4 (VSA4) isolated from Ehrlichia chaffeensis, which

CC has similarity to major antigen protein (MAP). The MAP polynucleotides

CC and polypeptides are useful as vaccines for conferring immunity to

CC rickettsia infection, including Cowdria ruminantium causing heartwater.

CC The MAP polynucleotides may be used as molecular markers in nucleic acid

CC analysis procedures, and to produce the MAP polypeptides, which may

CC be used to raise antibodies that are reactive with the polypeptides.

CC The nucleic acids may further be used as probes to identify

CC complementary sequences within other nucleic acid molecules or genomes,

CC where such probes can be applied to identify or distinguish infectious

CC strains of organisms in diagnostic procedures or in rickettsial

CC research where identification of particular organisms or strains is

CC needed.

XX CC

XX SQ Sequence 280 AA;

Query Match 63.9%; Score 62; DB 22; Length 280;

Best Local Similarity 60.0%; Pred. No. 0.018;

Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15

DB 60 ntgtgvgfkgdwdg 74

II :|||||I

RESULT 9

AA06948

ID AAY06948 standard; Protein; 280 AA.

XX AC AAY06948;

XX DT 05-JUL-1999 (first entry)

XX DE E. chaffeensis OMP-1F protein.

XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;

XX KX detection; dog.

XX OS Ehrlichia chaffeensis.

XX PN WO9913720-A1.

XX PD 25-MAR-1999.

XX PF 18-SEP-1998; 98WO-US19600.

XX PR 19-SEP-1997; 97US-0059353.

XX PA (OHIS) UNIV OHIO STATE.

XX PI Ohashi N, Rikihisa Y;

XX PS Disclosure; Fig 8B; 55pp; English.

XX CC The invention provides isolated outer membrane proteins (OMP) from

CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part

CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown

CC in AAY06943-958. The E. canis proteins form part of the P30 family and

CC consist of proteins shown in AAY06959-970. The proteins and genes are

CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX CC

XX SQ Sequence 280 AA;

Query Match 60.8%; Score 59; DB 20; Length 280;

Best Local Similarity 60.0%; Pred. No. 0.057;

Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15

DB 60 ntgtgvgfkgdwdg 74

II :|||||I

RESULT 10

AA06959

ID AAY06959 standard; Protein; 288 AA.

XX AC AAY06959;

XX DT 05-JUL-1999 (first entry)

XX DE E. canis P30 protein.

XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;

XX KX detection; dog.

XX OS Ehrlichia canis.

XX PN WO9913720-A1.

XX PD 25-MAR-1999.

XX PF 18-SEP-1998; 98WO-US19600.

XX PR 19-SEP-1997; 97US-0059353.

XX PA (OHIS) UNIV OHIO STATE.

XX PI Ohashi N, Rikihisa Y;

XX PS WPI; 1999-254290/21.

XX DR N-PSDB; AAX34759.

XX XI Novel outer membrane proteins from Ehrlichia chaffeensis and

PT Ehrlichia canis

XX PS Disclosure; Fig 19B; 55pp; English.

XX CC The invention provides isolated outer membrane proteins (OMP) from

CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part

CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown

CC in AAY06943-958. The E. canis proteins form part of the P30 family and

CC consist of proteins shown in AAY06959-970. The proteins and genes are

CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX CC

XX SQ Sequence 288 AA;

Query Match 60.8%; Score 59; DB 20; Length 288;

Best Local Similarity 60.0%; Pred. No. 0.059;

Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15

DB 60 ntgtgvgfkgdwdg 74

II :|||||I

RESULT 11

AA06942

ID AAY06942 standard; Protein; 256 AA.

XX AC AAY06942;

XX DT 05-JUL-1999 (first entry)

XX

XX PS Example 3; Fig 2A-2B; 30pp; English.

XX CC The sequence represents the amino acid sequence of variable surface

CC antigen 4 (VSA4) isolated from Ehrlichia chaffeensis, which

CC has similarity to major antigen protein (MAP). The MAP polynucleotides

CC and polypeptides are useful as vaccines for conferring immunity to

CC rickettsia infection, including Cowdria ruminantium causing heartwater.

CC The MAP polynucleotides may be used as molecular markers in nucleic acid

CC analysis procedures, and to produce the MAP polypeptides, which may

CC be used to raise antibodies that are reactive with the polypeptides.

CC The nucleic acids may further be used as probes to identify

CC complementary sequences within other nucleic acid molecules or genomes,

CC where such probes can be applied to identify or distinguish infectious

CC strains of organisms in diagnostic procedures or in rickettsial

CC research where identification of particular organisms or strains is

CC needed.

XX CC

XX SQ Sequence 280 AA;

Query Match 60.8%; Score 59; DB 20; Length 280;

Best Local Similarity 60.0%; Pred. No. 0.057;

Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15

DB 60 ntgtgvgfkgdwdg 74

II :|||||I

RESULT 10

AA06959

ID AAY06959 standard; Protein; 288 AA.

XX AC AAY06959;

XX DT 05-JUL-1999 (first entry)

XX DE E. canis P30 protein.

XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;

XX KX detection; dog.

XX OS Ehrlichia canis.

XX PN WO9913720-A1.

XX PD 25-MAR-1999.

XX PF 18-SEP-1998; 98WO-US19600.

XX PR 19-SEP-1997; 97US-0059353.

XX PA (OHIS) UNIV OHIO STATE.

XX PI Ohashi N, Rikihisa Y;

XX PS WPI; 1999-254290/21.

XX DR N-PSDB; AAX34759.

XX XI Novel outer membrane proteins from Ehrlichia chaffeensis and

PT Ehrlichia canis

XX PS Disclosure; Fig 19B; 55pp; English.

XX CC The invention provides isolated outer membrane proteins (OMP) from

CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part

CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown

CC in AAY06943-958. The E. canis proteins form part of the P30 family and

CC consist of proteins shown in AAY06959-970. The proteins and genes are

CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX CC

XX SQ Sequence 288 AA;

Query Match 60.8%; Score 59; DB 20; Length 288;

Best Local Similarity 60.0%; Pred. No. 0.059;

Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15

DB 60 ntgtgvgfkgdwdg 74

II :|||||I

RESULT 11

AA06942

ID AAY06942 standard; Protein; 256 AA.

XX AC AAY06942;

XX DT 05-JUL-1999 (first entry)

XX

DE E. chaffeensis p28 protein.
 XX
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.
 XX
 XX Ehrlichia chaffeensis.
 XX
 XX WO9913720-A1.
 XX
 XX 25-MAR-1999.
 XX
 XX 18-SEP-1998; 98WO-US19600.
 XX
 XX 19-SEP-1997; 97US-0059353.
 XX
 XX (OHIS) UNIV OHIO STATE.
 XX
 XX Ohashi N, Rikihisa Y;
 XX WPI; 1999-254290/21.
 XX N-PSDB; AAX34742.
 XX
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis
 PT
 PS Claim 18; Fig 1; 55pp; English.
 XX
 XX The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX
 XX Sequence 256 AA;

Query Match 59.8%; Score 58; DB 20; Length 256;
 Best Local Similarity 60.0%; Pred No. 0.076; 2; Indels 0; Gaps 0;
 Matches 9; Conservative 4; Mismatches 2;

QY 1 NPTVALYGLKQWEG 15
 Db 34 nttvgvfgiknwdg 48

RESULT 12
 AAW51095
 ID AAW51095 standard; Protein; 276 AA.
 XX
 XX AAW51095;
 XX
 XX 14-SEP-1998 (first entry)
 XX
 XX Ehrlichia chaffeensis VSA5 protein (partial sequence).
 XX
 XX MAP1 homologue; variable surface antigen; VSA5; rickettsia;
 KW DNA vaccine.
 KW
 XX Ehrlichia chaffeensis.
 XX
 XX Key Location/Qualifiers
 FH Peptide
 FT 1..25
 FT /note= "putative signal peptide"
 XX
 XX WO9816554-A1.
 XX
 XX 23-APR-1998.
 XX
 XX 17-OCT-1997; 97WO-US19044.
 XX
 XX 17-OCT-1996; 96US-0733230.
 XX

PA (UYFL) UNIV FLORIDA.
 XX
 XX Barbet AF, Burrigide MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 XX
 XX WPI; 1998-251232/22.
 XX N-PSDB; AAV07179.
 XX
 XX Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX
 XX Claim 3; Fig 2B; 39pp; English.
 XX
 XX This is the near full-length variable surface antigen VSA5 protein
 CC of Ehrlichia chaffeensis; it lacks about 5-7 C-terminal amino acid
 CC residues. The VSA5 amino acid sequence was deduced from a partial
 CC open reading frame (ORF5) of a genomic locus (see AAV07179) of E.
 CC chaffeensis that was obtained on the basis of homology to the major
 CC antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium. This
 CC genomic locus included 5 ORFs encoding similar, but non-identical
 CC proteins (see AAW51091-95). A claimed composition comprises a
 CC nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99)
 CC that elicits a protective immune response against a rickettsial
 CC pathogen. The nucleic acid is used, in human or veterinary
 CC medicine, in vaccines to protect against Rickettsia, Ehrlichia,
 CC Anaplasma and Cowdria species. The Ehrlichia antigenic
 CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).
 XX
 XX Sequence 276 AA;

Query Match 59.8%; Score 58; DB 19; Length 276;
 Best Local Similarity 60.0%; Pred No. 0.083; 2; Indels 0; Gaps 0;
 Matches 9; Conservative 4; Mismatches 2;

QY 1 NPTVALYGLKQWEG 15
 Db 59 nttvgvfgiknwdg 73

RESULT 13
 AAB36189
 ID AAB36189 standard; Protein; 276 AA.
 XX
 XX AAB36189;
 XX
 XX 02-MAR-2001 (first entry)
 XX
 XX Ehrlichia chaffeensis partial VSA5.
 XX
 XX Ehrlichia chaffeensis; VSA5; variable surface antigen 5; MAP1;
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
 KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
 KW 3gdorf3.
 XX
 XX Ehrlichia chaffeensis.
 OS
 XX WO200065063-A2.
 XX
 XX 02-NOV-2000.
 XX
 XX 21-APR-2000; 2000WO-US10886.
 XX
 XX 22-APR-1999; 99US-0130725.
 XX
 XX (UYFL) UNIV FLORIDA.
 XX
 XX Barbet AF, Bowie MV, Ganta RR, Burrigide MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Allemen AR;
 XX WPI; 2000-679675/66.
 XX

AA
CC
CC
CC

The sequence represents the amino acid sequence of variable surface antigen 5 (VSA5) isolated from *Ehrlichia chaffeensis*, which

CC 19 mcn. The Ehrlichia antigenic polypeptides can be used
CC diagnostically to detect antibodies associated with Ehrlichia
CC infection (claimed).

Thu Mar 14 10:13:46 2002

us-09-765-739a-6.rag

Page 8

XX
SQ

Sequence 280 AA;

Query Match 59.8%; Score 58; DB 19; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.084;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
| | : | | | : | |
Db 60 nttvgvfglkqwdg 74

Search completed: March 14, 2002, 09:09:38
Job time: 429 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:24:10 ; Search time 103.37 seconds
(without alignments)
26.886 Million cell updates/sec

Title: US-09-765-739A-7
Perfect score: 103
Sequence: 1 NTTTGVFLKQDWDGSTIS 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	103	100.0	280	2	052107 ehrlichia c
2	97	94.2	280	2	085357 ehrlichia c
3	96	93.2	288	2	092GJ2 ehrlichia c
4	89	86.4	246	2	09RH35 ehrlichia c
5	89	86.4	276	2	085817 ehrlichia c
6	88	85.4	276	2	085358 ehrlichia c
7	88	85.4	280	2	085816 ehrlichia c
8	88	85.4	280	2	09ZGM9 ehrlichia c
9	88	85.4	281	2	09ACI9 ehrlichia c
10	80	77.7	280	2	09P473 ehrlichia c
11	80	77.7	280	2	09ADV3 ehrlichia c
12	74	71.8	286	2	052105 ehrlichia c
13	73	70.9	278	2	09R8A9 ehrlichia c
14	73	70.9	278	2	09R8A8 ehrlichia c
15	73	70.9	278	2	09R8A7 ehrlichia c
16	73	70.9	278	2	09R8A6 ehrlichia c
17	73	70.9	278	2	09R8A5 ehrlichia c
18	73	70.9	278	2	09R3J3 ehrlichia c
19	73	70.9	278	2	09F472 ehrlichia c

20	73	70.9	307	2	Q9ZGJ1	Q9ZGJ1 ehrlichia c
21	72	69.9	276	2	Q9F475	Q9F475 ehrlichia c
22	60	58.3	278	2	052106	052106 ehrlichia c
23	60	58.3	280	2	052104	052104 ehrlichia c
24	59.5	57.8	265	2	Q9AF99	Q9AF99 cowdria rum
25	59.5	57.8	270	2	Q9AF98	Q9AF98 cowdria rum
26	59.5	57.8	287	2	Q46329	Q46329 cowdria rum
27	59.5	57.8	287	2	Q46331	Q46331 cowdria rum
28	59.5	57.8	287	2	Q9R425	Q9R425 cowdria rum
29	59.5	57.8	290	2	Q46324	Q46324 cowdria rum
30	59.5	57.8	290	2	Q46330	Q46330 cowdria rum
31	59.5	57.8	290	2	Q46332	Q46332 cowdria rum
32	59.5	57.8	290	2	Q46333	Q46333 cowdria rum
33	59	57.3	272	2	Q9AMF6	Q9AMF6 cowdria sp.
34	59	57.3	284	2	Q46327	Q46327 cowdria rum
35	59	57.3	284	2	Q9AF41	Q9AF41 cowdria rum
36	58	56.3	290	2	Q9AEU3	Q9AEU3 cowdria rum
37	56	54.4	281	2	Q46328	Q46328 cowdria rum
38	56	54.4	281	2	Q9S6H1	Q9S6H1 cowdria rum
39	56	54.4	281	2	Q9S6H0	Q9S6H0 cowdria rum
40	55	53.4	133	2	085360	085360 ehrlichia c
41	55	53.4	283	2	Q9F474	Q9F474 ehrlichia c
42	52.5	51.0	271	2	Q9AF40	Q9AF40 cowdria rum
43	52	50.5	435	10	Q9FKI3	Q9FKI3 arabidopsis
44	47	45.6	584	2	Q06352	Q06352 escherichia
45	45	43.7	144	5	Q9U993	Q9U993 acanthocheil

ALIGNMENTS

RESULT 1					
052107					
ID	052107	PRELIMINARY;	PRT;	280	AA.
AC	052107;				
DT	01-JUN-1998 (TREMBLrel. 06, Created)				
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)				
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)				
DE	OMP-1F.				
GN	OMP-1F.				
OS	Ehrlichia chaffeensis.				
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;				
OC	Rickettsiaceae; Ehrlichieae; Ehrlichia.				
OX	NCBI_TaxID=945;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ARKANSAS;				
RX	MEDLINE-98084465; PubMed-9423849;				
RA	Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;				
RT	"Immunodominant major outer membrane proteins of Ehrlichia chaffeensis are encoded by a polymorphic multigene family."				
RL	Infect Immun. 66:132-139(1998).				
DR	EMBL; U72231; AAC02940.1; .				
DR	InterPro; IPR002566; Surface_Ag_msp4.				
DR	Pfam; PF01617; Surface_Ag_2; 1.				
SQ	SEQUENCE 280 AA; 30731 MW; CCAA6C34E2AF393E CRG64;				

Query Match 100.0%; Score 103; DB 2; Length 280;
Best Local Similarity 100.0%; Pred. No. 8.6e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	NTTTGVFLKQDWDGSTIS 19
Db	60	NTTTGVFLKQDWDGSTIS 78

RESULT 2					
ID	085357	PRELIMINARY;	PRT;	280	AA.
AC	085357;				
DT	01-NOV-1998 (TREMBLrel. 08, Created)				
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)				

```

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 28 KDA MAJOR SURFACE ANTIGEN-4.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Allemen A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichieae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL: AF062761; AAC26720.1;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2.1.
DR SEQUENCE 280 AA; 30743 MW; FBB841DAF08EE4DC CRC64;

Query Match 94.2%; Score 97; DB 2; Length 280;
Best Local Similarity 94.7%; Pred. No. 8.1e-08;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGSTIS 19
DB 60 NTTTGVFGLKQDWGSTIS 78

RESULT 3
ID Q9ZGJ2 PRELIMINARY; PRT; 288 AA.
AC Q9ZGJ2;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 30-KDA MAJOR OUTER MEMBRANE PROTEIN (P28-8).
DE P30 OR P28-8.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
DR EMBL: AF078553; AAC68667.1;
DR EMBL: AF082744; AAG14362.1;
DR InterPro: IPR002566; Surface_Ag_msp4.

DR Pfam: PF01617; Surface_Ag_2.1.
DR SEQUENCE 276 AA; 30027 MW; 2FD3698FCF1F60BE CRC64;

DR Pfam: PF01617; Surface_Ag_2.1.
DR SEQUENCE 288 AA; 31590 MW; 86DCAEC88E9BF5E CRC64;

Query Match 93.2%; Score 96; DB 2; Length 288;
Best Local Similarity 94.4%; Pred. No. 1.2e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGSTI 18
DB 60 NTTTGVFGLKQDWGSTI 77

RESULT 4
ID Q9RH35 PRELIMINARY; PRT; 246 AA.
AC Q9RH35;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR (FRAGMENT).
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ST. VINCENT;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
DR EMBL: AF077735; AAC31548.1;
DR InterPro: IPR001702; Gram_neg_porin.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2.1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR NON_TER
FT SEQUENCE 246 AA; 26884 MW; C9776392C5129A2F CRC64;

Query Match 86.4%; Score 89; DB 2; Length 246;
Best Local Similarity 88.9%; Pred. No. 1.4e-06;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGSTI 18
DB 29 NTTAGVFGLKQDWGSAI 46

RESULT 5
ID O85817 PRELIMINARY; PRT; 276 AA.
AC O85817;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAPULPA;
RA Yu X.J., Walker D.H.;
RT "Ehrlichia chaffeensis 28 kDa outer membrane protein.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF077734; AAC31547.1;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2.1.
DR SEQUENCE 276 AA; 30027 MW; 2FD3698FCF1F60BE CRC64;
```

Query Match 85.48; Score 88; DB 2; Length 280;

DR EMBL; U72291; AAK28673.1; -.
SQ SEQUENCE 281 AA: 30243 MW:

Query Match 85.48; Score 88; DB 2; Length 280;

```
Query Match      85.4%; Score 88; DB 2; Length 281;
Best Local Similarity 84.2%; Pred. No. 2.3e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGSTS 19
   ||| ||||| ||||| ||
Db 59 NTTGVFGLKQWDGSAIS 77

RESULT 10
Q9F473 PRELIMINARY: PRT; 280 AA.
AC Q9F473;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE P28-6.
GN P28-6.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RA MEDLINE=99242757; PubMed=10225842;
RX McBride J.W., Yu, X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RA MEDLINE=20432107; PubMed=10974556;
RX McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
DR EMBL; AF082744; AAG14361.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30762 MW; BE284A4B94FE3123 CRC64;

Query Match      77.7%; Score 80; DB 2; Length 280;
Best Local Similarity 73.7%; Pred. No. 4.6e-05;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGSTS 19
   ||| ||||| ||||| ||
Db 59 NSTGVFGLKHDWNGTIS 77

RESULT 11
Q9ADV3 PRELIMINARY: PRT; 280 AA.
AC Q9ADV3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN P30-2.
GN P30-2.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;

Query Match      71.8%; Score 74; DB 2; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00044;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGSTS 19
   ||| ||||| ||||| ||
Db 60 NTTGVFGLKQDWDRCVIS 78
```

```
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF078553; AAK28699.1; -.
SQ SEQUENCE 280 AA; 30803 MW; 27238BE1C7E68A91 CRC64;

Query Match      77.7%; Score 80; DB 2; Length 280;
Best Local Similarity 73.7%; Pred. No. 4.6e-05;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGSTS 19
   ||| ||||| ||||| ||
Db 59 NSTGVFGLKHDWNGTIS 77

RESULT 12
Q52105 PRELIMINARY: PRT; 286 AA.
AC Q52105;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE OMP-1D.
GN OMP-1D.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., BurrIDGE M.J.,
RA Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichieae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL; U72291; AAC02938.1; -.
DR EMBL; AF062761; AAC26718.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 286 AA; 31509 MW; F145A79270F386BE CRC64;

Query Match      71.8%; Score 74; DB 2; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00044;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGSTS 19
   ||| ||||| ||||| ||
Db 60 NTTGVFGLKQDWDRCVIS 78
```


RESULT 15

Q9R8A7
ID Q9R8A7 PRELIMINARY; PRT; 278 AA.
AC Q9R8A7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE 28 KDA OUTER MEMBRANE PROTEIN (FRAGMENT).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DEMON (NORTH CAROLINA);
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of a conserved major immunoreactive 28-kilodalton
RT protein gene from a polymorphic multiple gene family of Ehrlichia
RT canis.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF082745; AAC64551.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 70.9%; Score 73; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.00062;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 2 TTGTGVLKQDWDGSKI 18
: | | | | | | | | | |
Db 61 STGVFGLKHDWDGSKI 77

Search completed: March 14, 2002, 09:24:10
Job time: 981 sec

RESULT 13

Q9R8A9
ID Q9R8A9 PRELIMINARY; PRT; 278 AA.
AC Q9R8A9;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE 28 KDA OUTER MEMBRANE PROTEIN (FRAGMENT).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LOUISIANA;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of a conserved major immunoreactive 28-kilodalton
RT protein gene from a polymorphic multiple gene family of Ehrlichia
RT canis.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF082745; AAC64551.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 70.9%; Score 73; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.00062;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 2 TTGTGVLKQDWDGSKI 18
: | | | | | | | | | |
Db 61 STGVFGLKHDWDGSKI 77

RESULT 14

Q9R8A8
ID Q9R8A8 PRELIMINARY; PRT; 278 AA.
AC Q9R8A8;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE 28 KDA OUTER MEMBRANE PROTEIN (FRAGMENT).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of a conserved major immunoreactive 28-kilodalton
RT protein gene from a polymorphic multiple gene family of Ehrlichia
RT canis.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF082746; AAC64552.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 70.9%; Score 73; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.00062;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 2 TTGTGVLKQDWDGSKI 18
: | | | | | | | | | |
Db 61 STGVFGLKHDWDGSKI 77

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:07:42 ; Search time 54.54 seconds
(without alignments)
7.839 Million cell updates/sec

Title: US-09-765-739A-7
Perfect score: 103
Sequence: 1 NTTTGVFLKQDWGSGTIS 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97	94.2	280	4	US-08-953-326-17
2	88	85.4	276	4	US-08-953-326-18
3	88	85.4	280	3	US-08-733-230-4
4	88	85.4	280	4	US-08-953-326-4
5	74	71.8	286	4	US-08-953-326-15
6	60	58.3	278	4	US-08-953-326-16
7	59.5	57.8	287	3	US-08-733-230-2
8	59.5	57.8	287	4	US-08-953-326-2
9	55	53.4	133	4	US-08-953-326-20
10	41	39.8	135	1	US-08-446-600A-4
11	40	38.8	1025	3	US-08-304-309-2
12	40	38.8	1025	3	US-08-991-942-2
13	40	38.8	1025	4	US-09-138-103-2
14	40	38.8	1025	5	PCT-US95-04567-4
15	39	37.9	1025	2	US-08-304-309-4
16	39	37.9	1025	3	US-08-991-942-4
17	39	37.9	1025	5	PCT-US95-04567-2
18	38	36.9	371	3	US-08-586-165-3
19	38	36.9	372	3	US-08-586-165-5
20	38	36.9	859	1	US-08-053-614-2
21	38	36.9	859	1	US-08-316-397B-2
22	38	36.9	859	2	US-09-034-306-2
23	38	36.9	859	4	US-09-259-437-2
24	38	36.9	859	5	PCT-US93-09782-2
25	38	36.9	1181	1	US-08-053-614-4
26	38	36.9	1181	1	US-08-316-397B-4
27	38	36.9	1181	2	US-09-034-306-4

28	38	36.9	1181	4	US-09-259-437-4	Sequence 4, Appli
29	38	36.9	1181	5	PCT-US93-09782-4	Sequence 4, Appli
30	38	36.9	1255	3	US-08-947-823-3	Sequence 3, Appli
31	38	36.9	1257	3	US-08-947-823-5	Sequence 5, Appli
32	37	35.9	31	1	US-08-190-802A-70	Sequence 70, Appl
33	37	35.9	31	1	US-08-190-802A-129	Sequence 129, App
34	37	35.9	31	1	US-08-190-802A-173	Sequence 173, App
35	37	35.9	31	4	US-08-477-346-70	Sequence 70, Appl
36	37	35.9	31	4	US-08-477-346-129	Sequence 129, App
37	37	35.9	31	4	US-08-477-346-173	Sequence 173, App
38	37	35.9	164	4	US-09-020-956-178	Sequence 178, App
39	37	35.9	164	4	US-09-030-607-178	Sequence 178, App
40	37	35.9	317	1	US-08-190-802A-27	Sequence 27, Appl
41	37	35.9	317	1	US-08-190-802A-41	Sequence 41, Appl
42	37	35.9	317	1	US-08-190-802A-47	Sequence 47, Appl
43	37	35.9	317	4	US-08-477-346-27	Sequence 27, Appl
44	37	35.9	317	4	US-08-477-346-41	Sequence 41, Appl
45	37	35.9	317	4	US-08-477-346-47	Sequence 47, Appl

ALIGNMENTS

RESULT 1
US-08-953-326-17
Sequence 17, Application US/08953326
Patent No. 6251872

GENERAL INFORMATION:
APPLICANT: Barbet, Anthony F.
APPLICANT: Ganta, Roman R.
APPLICANT: McGuire, Travis C.
APPLICANT: Burridge, Michael J.
APPLICANT: Nyika, Aceme
APPLICANT: Rurangirwa, Fred R.
APPLICANT: Mahan, Suman M.

TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
Animals and Humans
FILE REFERENCE: UF-167C1
CURRENT APPLICATION NUMBER: US/08/953,326
CURRENT FILING DATE: 1997-10-17
EARLIER APPLICATION NUMBER: 08/953,326
EARLIER FILING DATE: 1997-10-17
EARLIER APPLICATION NUMBER: 08/733,230
EARLIER FILING DATE: 1996-10-17
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 17
LENGTH: 280
TYPE: PRT
ORGANISM: Ehrlichia chaffeensis
US-08-953-326-17

Query Match 94.2%; Score 97; DB 4; Length 280;
Best Local Similarity 94.7%; Pred. No. 1.4e-08;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NTTTGVFLKQDWGSGTIS 19
||| ||||| ||||| |||||
Db 60 NTTTGVFLKQDWGSGTIS 78

RESULT 2
US-08-953-326-18
Sequence 18, Application US/08953326
Patent No. 6251872
GENERAL INFORMATION:
APPLICANT: Barbet, Anthony F.
APPLICANT: Ganta, Roman R.
APPLICANT: McGuire, Travis C.
APPLICANT: Burridge, Michael J.
APPLICANT: Nyika, Aceme
APPLICANT: Rurangirwa, Fred R.

```
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-18

Query Match      85.4%; Score 88; DB 4; Length 276;
Best Local Similarity 84.2%; Pred. No. 4.3e-07;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWDGSTIT 19
   ||| ||||| ||||| ||
Db 59 NNTTGVFGLKQDWDGSAIS 77

RESULT 3
US-08-733-230-4
; Sequence 4, Application US/08/733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; FILE OF INVENTION: Rickettsial Diseases and Methods of Use
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-4
```

```
Query Match      85.4%; Score 88; DB 3; Length 280;
Best Local Similarity 84.2%; Pred. No. 4.3e-07;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWDGSTIT 19
   ||| ||||| ||||| ||
Db 60 NNTTGVFGLKQDWDGSAIS 78

RESULT 4
US-08-953-326-4
; Sequence 4, Application US/089533326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4

Query Match      85.4%; Score 88; DB 4; Length 280;
Best Local Similarity 84.2%; Pred. No. 4.3e-07;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWDGSTIT 19
   ||| ||||| ||||| ||
Db 60 NNTTGVFGLKQDWDGSAIS 78

RESULT 5
US-08-953-326-15
; Sequence 15, Application US/089533326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
```

; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-15

Query Match 71.8%; Score 74; DB 4; Length 286;
Best Local Similarity 68.4%; Pred. No. 9.5e-05;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NNTTGVFLKQDWDGSGTIS 19
||| ||||:|||| ||
DB 60 NRTGVFGIEQDWDRCVIS 78

RESULT 6
US-08-953-326-16
; Sequence 16, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 16
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-16

Query Match 58.3%; Score 60; DB 4; Length 278;
Best Local Similarity 52.6%; Pred. No. 0.019;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NNTTGVFLKQDWDGSGTIS 19
||| ||||:|||| ||
DB 60 NPTVALYGLKQDWEIGISS 78

RESULT 7
US-08-733-230-2
; Sequence 2, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-2

Query Match 57.8%; Score 59.5; DB 3; Length 287;
Best Local Similarity 80.0%; Pred. No. 0.024;
Matches 12; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 1 NNTTGVFLKQDWDG 15
||| ||||:|||| ||
DB 62 NTQT-VFGLKQDWDG 75

RESULT 8
US-08-953-326-2
; Sequence 2, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
US-08-953-326-2

Query Match 57.8%; Score 59.5; DB 4; Length 287;
Best Local Similarity 80.0%; Pred. No. 0.024;
Matches 12; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 1 NNTTGVFLKQDWDG 15

```
Db 62 NTQT-VFGLKQWDG 75
      || | |||||:||||
NAME: Ann G. Robinson
REGISTRATION NUMBER: 39,820
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (513) 651-6128
TELEFAX: (513) 651-6981
TELEX: 21-4396 F&J Cin
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 135
TYPE: Amino Acid
TOPOLOGY: Linear
MOLECULE TYPE: Protein
US-08-446-600A-4

Query Match 39.8%; Score 41; DB 1; Length 135;
Best Local Similarity 47.6%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 NTTGVFGLKQWDG--STIS 19
      || | |||||:||||
Db 88 NETDGALVQHQMKGKSTIT 108
      || | |||||:||||

RESULT 11
US-08-304-309-2
: Sequence 2, Application US/08304309
: Patent No. 5856454
: GENERAL INFORMATION:
: APPLICANT: GONZALEZ, Frank J.
: APPLICANT: FERNANDEZ-SALGUERO, Pedro
: TITLE OF INVENTION: CLONING AND EXPRESSION OF CDNA FOR HUMAN
: TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourie and Crew
: STREET: Steuart Street tower, One Market Plaza
: CITY: San Francisco
: STATE: California
: COUNTRY: US
: ZIP: 94105-1493
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/304,309
: FILING DATE: 09-SEP-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Smith, Timothy L.
: REGISTRATION NUMBER: 35,367
: REFERENCE/DOCKET NUMBER: 15280-210
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1025 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-304-309-2

Query Match 38.8%; Score 40; DB 2; Length 1025;
Best Local Similarity 58.3%; Pred. No. 1.9e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQD 12
      || | |||||:||||
```

```
Db 62 NTQT-VFGLKQWDG 75
      || | |||||:||||
NAME: Ann G. Robinson
REGISTRATION NUMBER: 39,820
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (513) 651-6128
TELEFAX: (513) 651-6981
TELEX: 21-4396 F&J Cin
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 135
TYPE: Amino Acid
TOPOLOGY: Linear
MOLECULE TYPE: Protein
US-08-446-600A-4

Query Match 39.8%; Score 41; DB 1; Length 135;
Best Local Similarity 47.6%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 NTTGVFGLKQWDG--STIS 19
      || | |||||:||||
Db 88 NETDGALVQHQMKGKSTIT 108
      || | |||||:||||

RESULT 11
US-08-304-309-2
: Sequence 2, Application US/08304309
: Patent No. 5856454
: GENERAL INFORMATION:
: APPLICANT: GONZALEZ, Frank J.
: APPLICANT: FERNANDEZ-SALGUERO, Pedro
: TITLE OF INVENTION: CLONING AND EXPRESSION OF CDNA FOR HUMAN
: TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourie and Crew
: STREET: Steuart Street tower, One Market Plaza
: CITY: San Francisco
: STATE: California
: COUNTRY: US
: ZIP: 94105-1493
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/304,309
: FILING DATE: 09-SEP-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Smith, Timothy L.
: REGISTRATION NUMBER: 35,367
: REFERENCE/DOCKET NUMBER: 15280-210
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1025 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-304-309-2

Query Match 38.8%; Score 40; DB 2; Length 1025;
Best Local Similarity 58.3%; Pred. No. 1.9e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQD 12
      || | |||||:||||
```

```
Db 62 NTQT-VFGLKQWDG 75
      || | |||||:||||
NAME: Ann G. Robinson
REGISTRATION NUMBER: 39,820
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (513) 651-6128
TELEFAX: (513) 651-6981
TELEX: 21-4396 F&J Cin
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 135
TYPE: Amino Acid
TOPOLOGY: Linear
MOLECULE TYPE: Protein
US-08-446-600A-4

Query Match 39.8%; Score 41; DB 1; Length 135;
Best Local Similarity 47.6%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 NTTGVFGLKQWDG--STIS 19
      || | |||||:||||
Db 88 NETDGALVQHQMKGKSTIT 108
      || | |||||:||||

RESULT 11
US-08-304-309-2
: Sequence 2, Application US/08304309
: Patent No. 5856454
: GENERAL INFORMATION:
: APPLICANT: GONZALEZ, Frank J.
: APPLICANT: FERNANDEZ-SALGUERO, Pedro
: TITLE OF INVENTION: CLONING AND EXPRESSION OF CDNA FOR HUMAN
: TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourie and Crew
: STREET: Steuart Street tower, One Market Plaza
: CITY: San Francisco
: STATE: California
: COUNTRY: US
: ZIP: 94105-1493
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/304,309
: FILING DATE: 09-SEP-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Smith, Timothy L.
: REGISTRATION NUMBER: 35,367
: REFERENCE/DOCKET NUMBER: 15280-210
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1025 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-304-309-2

Query Match 38.8%; Score 40; DB 2; Length 1025;
Best Local Similarity 58.3%; Pred. No. 1.9e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQD 12
      || | |||||:||||
```

```
Db 62 NTQT-VFGLKQWDG 75
      || | |||||:||||
NAME: Ann G. Robinson
REGISTRATION NUMBER: 39,820
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (513) 651-6128
TELEFAX: (513) 651-6981
TELEX: 21-4396 F&J Cin
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 135
TYPE: Amino Acid
TOPOLOGY: Linear
MOLECULE TYPE: Protein
US-08-446-600A-4

Query Match 39.8%; Score 41; DB 1; Length 135;
Best Local Similarity 47.6%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 NTTGVFGLKQWDG--STIS 19
      || | |||||:||||
Db 88 NETDGALVQHQMKGKSTIT 108
      || | |||||:||||

RESULT 11
US-08-304-309-2
: Sequence 2, Application US/08304309
: Patent No. 5856454
: GENERAL INFORMATION:
: APPLICANT: GONZALEZ, Frank J.
: APPLICANT: FERNANDEZ-SALGUERO, Pedro
: TITLE OF INVENTION: CLONING AND EXPRESSION OF CDNA FOR HUMAN
: TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourie and Crew
: STREET: Steuart Street tower, One Market Plaza
: CITY: San Francisco
: STATE: California
: COUNTRY: US
: ZIP: 94105-1493
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/304,309
: FILING DATE: 09-SEP-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Smith, Timothy L.
: REGISTRATION NUMBER: 35,367
: REFERENCE/DOCKET NUMBER: 15280-210
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1025 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-304-309-2

Query Match 38.8%; Score 40; DB 2; Length 1025;
Best Local Similarity 58.3%; Pred. No. 1.9e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQD 12
      || | |||||:||||
```

Db 736 NTVSGLMGLKSD 747

RESULT 12

US-08-991-942-2
 ; Sequence 2, Application US/08991942
 ; Patent No. 6015673
 ; GENERAL INFORMATION:
 ; APPLICANT: GONZALEZ, Frank J.
 ; APPLICANT: FERNANDEZ-SALGUERO, Pedro
 ; TITLE OF INVENTION: CLONING AND EXPRESSION OF CDNA FOR HUMAN
 ; TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE
 ; NUMBER OF SEQUENCES: 13
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend Khourie and Crew
 ; STREET: Stewart Street Tower, One Market Plaza
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: US
 ; ZIP: 94105-1493
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/991,942
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/304,309
 ; FILING DATE: 09-SEP-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Smith, Timothy L.
 ; REGISTRATION NUMBER: 35,367
 ; REFERENCE/DOCKET NUMBER: 15280-210
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 543-9600
 ; TELEFAX: (415) 543-5043
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1025 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-991-942-2

Query Match 38.8%; Score 40; DB 3; Length 1025;

Best Local Similarity 58.3%; Pred. No. 1.9e+02;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NNTTGVFGLKQD 12

||:|:||||

Db 736 NTVSGLMGLKSD 747

RESULT 13

US-09-138-103-2
 ; Sequence 2, Application US/09138103A
 ; Patent No. 6232448
 ; GENERAL INFORMATION:
 ; APPLICANT: Yoshikubo, Takashi
 ; APPLICANT: Hasegawa, Masami
 ; TITLE OF INVENTION: Immunological Materials and Methods for Detecting
 ; TITLE OF INVENTION: Dihydropyrimidine Dehydrogenase
 ; FILE REFERENCE: 09/138,103 Yoshikubo, et al.
 ; CURRENT APPLICATION NUMBER: US/09/138,103A
 ; CURRENT FILING DATE: 1998-08-21
 ; EARLIER APPLICATION NUMBER: 97114630.3
 ; EARLIER FILING DATE: 1997-08-22
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2
 ; LENGTH: 1025
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-138-103-2

Query Match 38.8%; Score 40; DB 4; Length 1025;

Best Local Similarity 58.3%; Pred. No. 1.9e+02;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NNTTGVFGLKQD 12

||:|:||||

Db 736 NTVSGLMGLKSD 747

RESULT 14

PCT-US95-04567-4
 ; Sequence 4, Application PC/TUS9504567
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE
 ; NUMBER OF SEQUENCES: 28
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P.O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: United States of America
 ; ZIP: 77210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
 ; SOFTWARE: PatentIn Release #1.0, Version
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/04567
 ; FILING DATE: CONCURRENTLY HERewith
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/227,357
 ; FILING DATE: 13-APR-1994
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Wilson, Mark B.
 ; REGISTRATION NUMBER: 37,259
 ; REFERENCE/DOCKET NUMBER: UOAB025P--
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (512) 418-3000
 ; TELEFAX: (713) 789-2679
 ; TELEX: 79-0924
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1025 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; PCT-US95-04567-4

Query Match 38.8%; Score 40; DB 5; Length 1025;

Best Local Similarity 58.3%; Pred. No. 1.9e+02;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NNTTGVFGLKQD 12

||:|:||||

Db 736 NTVSGLMGLKSD 747

RESULT 15

US-08-304-309-4

```

? Sequence 4, Application US/08304309
? Patent No. 5856454
? GENERAL INFORMATION:
? APPLICANT: GONZALEZ, Frank J.
? APPLICANT: FERNANDEZ-SALGUERO, Pedro
? TITLE OF INVENTION: CLONING AND EXPRESSION OF cDNA FOR HUMAN
? TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE
? NUMBER OF SEQUENCES: 13
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Townsend and Townsend Kourie and Crew
? STREET: Stewart Street Tower, One Market Plaza
? CITY: San Francisco
? STATE: California
? COUNTRY: US
? ZIP: 94105-1493
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/304,309
? FILING DATE: 09-SEP-1994
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: Smith, Timothy L.
? REGISTRATION NUMBER: 35,367
? REFERENCE/DOCKET NUMBER: 15280-210
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 543-9600
? TELEFAX: (415) 543-5043
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1025 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-304-309-4

```

```

Query Match      37.9%; Score 39; DB 2; Length 1025;
Best Local Similarity 58.3%; Pred. No. 2.8e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 NTWTGVFGLKQD 12
Db 736 NTVSGLMGLRAD 747

```

```

Search completed: March 14, 2002, 09:07:42
Job time: 843 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:09:38 ; Search time 111.55 Seconds
(without alignments)
12.617 Million cell updates/sec

Title: US-09-765-739A-7

Perfect score: 103

Sequence: 1 NNTTGVFGLKQDWDGSTIS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A_Geneseq_1101.*
- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103	100.0	280	20 AAY06948	E. chafeensis OMP-
2	97	94.2	280	19 AAW51094	Ehrlichia chaffeen
3	97	94.2	280	21 AAB36188	Ehrlichia chaffeen
4	97	94.2	280	22 AAU04198	Variable surface a
5	96	93.2	288	20 AAY06959	E. canis P30 prote
6	88	85.4	256	20 AAY06942	E. chafeensis p28
7	88	85.4	276	19 AAW51095	Ehrlichia chaffeen
8	88	85.4	276	21 AAB36189	Ehrlichia chaffeen
9	88	85.4	276	22 AAU04199	Variable surface a
10	88	85.4	280	19 AAW51089	Ehrlichia chaffeen
11	88	85.4	280	21 AAB36183	Ehrlichia chaffeen

12	88	85.4	280	22 AAU04193	Major antigenic pr
13	88	85.4	281	20 AAY06943	E. chafeensis OMP-
14	80	77.7	280	20 AAY06962	E. canis P30-2 pro
15	80	77.7	280	21 AAY71479	Ehrlichia canis im
16	74	71.8	286	19 AAW51092	Ehrlichia chaffeen
17	74	71.8	286	20 AAY06946	E. chafeensis OMP-
18	74	71.8	286	21 AAB36186	Ehrlichia chaffeen
19	74	71.8	286	22 AAU04196	Variable surface a
20	73	70.9	278	21 AAY71477	Ehrlichia canis im
21	73	70.9	307	20 AAY06961	E. canis P30-1 pro
22	72	69.9	276	20 AAY06964	Ehrlichia chaffeen
23	60	58.3	278	19 AAW51093	E. chafeensis OMP-
24	60	58.3	278	20 AAY06947	Ehrlichia chaffeen
25	60	58.3	278	21 AAB36187	Variable surface a
26	60	58.3	278	22 AAU04197	E. chafeensis OMP-
27	60	58.3	280	20 AAY06945	Cowdria ruminantiu
28	59.5	57.8	287	19 AAW51088	Cowdria ruminantiu
29	59.5	57.8	287	21 AAB36182	Major antigenic pr
30	59.5	57.8	287	22 AAU04192	Variable surface a
31	55	53.4	132	22 AAU04201	Ehrlichia canis VS
32	55	53.4	133	19 AAW51097	Ehrlichia canis pa
33	55	53.4	133	21 AAB36191	Ehrlichia canis im
34	55	53.4	133	21 AAY71480	Ehrlichia canis im
35	55	53.4	283	21 AAG71478	Ehrlichia canis im
36	52	50.5	240	21 AAG20226	Arabidopsis thalia
37	52	50.5	364	21 AAG47084	Arabidopsis thalia
38	52	50.5	365	21 AAG47111	Arabidopsis thalia
39	52	50.5	374	21 AAG47083	Arabidopsis thalia
40	52	50.5	375	21 AAG47110	Arabidopsis thalia
41	52	50.5	442	21 AAG47082	Arabidopsis thalia
42	52	50.5	443	21 AAG47109	Arabidopsis thalia
43	43	41.7	806	22 AAG92136	C glutamicum prote
44	43	41.7	904	20 AAY21976	Senescence-associa
45	41	39.8	135	15 AAR55866	Melanogenic inhibi

ALIGNMENTS

RESULT 1
AAY06948
ID AAY06948 standard; Protein; 280 AA.

AC AAY06948;
DT 05-JUL-1999 (first entry)
DE E. chafeensis OMP-1F protein.
KW Outer membrane protein; OMP; Ehrlichia chafeensis; E. canis; P30;
KW detection; dog.
OS Ehrlichia chafeensis.
PN WO9913720-A1.
PD 25-MAR-1999.
PF 18-SEP-1998; 98WO-US19600.
PR 19-SEP-1997; 97US-0059353.
PA (OHIS) UNIV OHIO STATE.
PI Ohashi N, Rikihisa Y;
DR WPI; 1999-254290/21.
DR N-PSDB; AAX34748.
PT Novel outer membrane proteins from Ehrlichia chafeensis and
PT Ehrlichia canis
PS Claim 16; Fig 8B; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the p30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX
 SQ Sequence 280 AA;

Query Match 100.0%; Score 103; DB 20; Length 280;
 Best Local Similarity 100.0%; Pred. No. 5.9e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NNTTGVFLKQDWDGSGTIS 19
 |||||
 Db 60 ntttgvflkqdwgdsgtis 78

RESULT 2
 AAW51094
 ID AAW51094 standard; Protein; 280 AA.
 XX
 AC AAW51094;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Ehrlichia chaffeensis VSA4 protein.
 XX
 KW MAP1 homologue; variable surface antigen; VSA4; rickettsia;
 KW DNA vaccine.
 XX
 OS Ehrlichia chaffeensis.
 XX
 FH Key Location/Qualifiers
 FT Peptide
 FT 1..25
 FT /note= "putative signal peptide"
 XX
 PN W09816554-A1.
 XX
 PD 23-APR-1998.
 XX
 PF 17-OCT-1997; 97WO-US19044.
 XX
 PR 17-OCT-1996; 96US-0733230.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 XX
 DR WPI: 1998-251232/22.
 DR N-PSDB; AAV07179.
 XX
 PT Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX
 PS Claim 3; Fig 2B; 39pp; English.
 XX
 CC This is the full-length variable surface antigen VSA4 protein of
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
 CC partial open reading frame (ORF4) of a genomic locus (see AAV07179)
 CC of E. chaffeensis that was obtained on the basis of homology to the
 CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
 CC This genomic locus included 5 ORFs encoding similar, but
 CC non-identical proteins (see AAW51091-95). A claimed composition
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
 CC (see AAW51088-99) that elicits a protective immune response against a
 CC rickettsial pathogen. The nucleic acid is used, in human or
 CC veterinary medicine, in vaccines to protect against Rickettsia,
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic

CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).
 XX
 SQ Sequence 280 AA;

Query Match 94.2%; Score 97; DB 19; Length 280;
 Best Local Similarity 94.7%; Pred. No. 5.7e-08;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 NNTTGVFLKQDWDGSGTIS 19
 |||||
 Db 60 ntttgvflkqdwgdsgtis 78

RESULT 3
 AAB36188
 ID AAB36188 standard; Protein; 280 AA.
 XX
 AC AAB36188;
 XX
 DT 02-MAR-2001 (first entry)
 XX
 DE Ehrlichia chaffeensis partial VSA4.
 XX
 KW Ehrlichia chaffeensis; VSA4; variable surface antigen 4: MAP1;
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
 KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
 KW 3gdorf3.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN W0200065063-A2.
 XX
 PD 02-NOV-2000.
 XX
 PF 21-APR-2000; 2000WO-US10886.
 XX
 PR 22-APR-1999; 99US-0130725.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Allemen AR;
 XX
 DR WPI: 2000-679675/66.
 DR N-PSDB; AAC68705.
 XX
 PT New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
 PT -
 XX
 PS Claim 3; Page 45-46; 63pp; English.
 XX
 CC The present sequence shows a high degree of similarity to the major
 CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
 CC used in a vaccines to protect animals or humans against rickettsial
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid vaccines
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
 CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.
 XX
 SQ Sequence 280 AA;

Query Match 94.2%; Score 97; DB 21; Length 280;
 Best Local Similarity 94.7%; Pred. No. 5.7e-08;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNTTGVFLKQDWGSGTIS 19
 ||| |||||||||||||
 Db 60 ntttgvgfllkqdwgsgtis 78

RESULT 4

AAU04198
 ID AAU04198 standard; Protein; 280 AA.

AC AAU04198;

DT 23-OCT-2001 (first entry)

DE Variable surface antigen 4 (VSA4) from Ehrlichia chaffeensis.

KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
 KW infection; heartwater; diagnostic; variable surface antigen; VSA.

OS Ehrlichia chaffeensis.

PN US6251872-B1.

PD 26-JUN-2001.

PF 17-OCT-1997; 97US-0953326.

PR 17-OCT-1996; 96US-0733230.

PA (UYFL) UNIV FLORIDA.

PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;

PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;

DR WPI; 2001-424487/45.

DR N-PSDB; AAS07578.

PT New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures -

PS Example 3; Fig 2A-2B; 30pp; English.

CC The sequence represents the amino acid sequence of variable surface
 CC antigen 4 (VSA4) isolated from Ehrlichia chaffeensis, which
 CC has similarity to major antigen protein (MAP). The MAP polynucleotides
 CC and polypeptides are useful as vaccines for conferring immunity to
 CC rickettsia infection, including Cowdria ruminantium causing heartwater.
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid
 CC analysis procedures, and to produce the MAP polypeptides, which may
 CC be used to raise antibodies that are reactive with the polypeptides.
 CC The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious
 CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is
 CC needed.

SQ Sequence 280 AA;

Query Match 94.2%; Score 97; DB 22; Length 280;
 Best Local Similarity 94.7%; Pred. No. 5.7e-08;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNTTGVFLKQDWGSGTIS 19
 ||| |||||||||||||
 Db 60 ntttgvgfllkqdwgsgtis 78

RESULT 5

AAU06959

ID AAU06959 standard; Protein; 288 AA.

AC AAU06959;

DT 05-JUL-1999 (first entry)

DE E. canis p30 protein.

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; p30;
 KW detection; dog.

OS Ehrlichia canis.

PN WO9913720-A1.

PD 25-MAR-1999.

PF 18-SEP-1998; 98WO-US19600.

PR 19-SEP-1997; 97US-0059353.

PA (OHIS) UNIV OHIO STATE.

PI Ohashi N, Rikihisa Y;

DR WPI; 1999-254290/21.

DR N-PSDB; AAX34759.

PT Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis

PS Disclosure; Fig 19B; 55pp; English.

CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAU06943-958. The E. canis proteins form part of the p30 family and
 CC consist of proteins shown in AAU06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.

SQ Sequence 288 AA;

Query Match 93.2%; Score 96; DB 20; Length 288;
 Best Local Similarity 94.4%; Pred. No. 8.6e-08;
 Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNTTGVFLKQDWGSGTIS 18
 ||||| |||||||||
 Db 60 ntttgvgfllkqdwgsgti 77

RESULT 6

AAU06942

ID AAU06942 standard; Protein; 256 AA.

AC AAU06942;

DT 05-JUL-1999 (first entry)

DE E. chaffeensis p28 protein.

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; p30;
 KW detection; dog.

OS Ehrlichia chaffeensis.

PN WO9913720-A1.

PD 25-MAR-1999.

PF 18-SEP-1998; 98WO-US19600.

PR 19-SEP-1997; 97US-0059353.

PA (OHIS) UNIV OHIO STATE.

PI Ohashi N, Rikihisa Y;

PI WPI: 1999-254290/21.

DR N-PSDB; AAX34742.

XX Novel outer membrane proteins from Ehrlichia chaffeensis and

PT Ehrlichia canis

XX Claim 18: Fig 1; 55pp: English.

XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the p30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 256 AA;

Query Match 85.4%; Score 88; DB 20; Length 256;

Best Local Similarity 84.2%; Pred. No. 1.5e-06;

Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWGDTIS 19

DB 34 nttvgvfglkqndgsais 52

RESULT 7

ID AAW51095 standard; Protein; 276 AA.

AC AAW51095;

DT 14-SEP-1998 (first entry)

DE Ehrlichia chaffeensis VSA5 protein (partial sequence).

XX MAP1 homologue; variable surface antigen; VSA5; rickettsia;

XX DNA vaccine.

XX Ehrlichia chaffeensis.

XX Key Location/Qualifiers

XX Peptide 1..25

XX /note= "putative signal peptide"

XX WO9816554-A1.

XX 23-APR-1998.

XX 17-OCT-1997; 97WO-US19044.

XX 17-OCT-1996; 96US-0733230.

XX (UYFL) UNIV FLORIDA.

XX Barbet AF, Burrig MJ, Ganta RR, Mahan SM, McGuire TC;

XX Nyika A, Rurangirwa FR;

XX WPI: 1998-251232/22.

XX N-PSDB; AAV07179.

XX Composition containing nucleic acid encoding rickettsial antigen -

XX useful for, e.g. stimulating protective immune response in humans or

XX animals

XX Claim 3; Fig 2b; 39pp; English.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

This is the near full-length variable surface antigen VSA5 protein of Ehrlichia chaffeensis; it lacks about 5-7 C-terminal amino acid residues. The VSA5 amino acid sequence was deduced from a partial open reading frame (ORF5) of a genomic locus (see AAV07179) of E. chaffeensis that was obtained on the basis of homology to the major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium. This genomic locus included 5 ORFs encoding similar, but non-identical proteins (see AAW51091-95). A claimed composition comprises a nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99) that elicits a protective immune response against a rickettsial pathogen. The nucleic acid is used, in human or veterinary medicine, in vaccines to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic polypeptides can be used diagnostically to detect antibodies associated with Ehrlichia infection (claimed).

Sequence 276 AA;

Query Match 85.4%; Score 88; DB 19; Length 276;

Best Local Similarity 84.2%; Pred. No. 1.7e-06;

Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWGDTIS 19

DB 59 nttvgvfglkqndgsais 77

RESULT 8

ID AAB36189 standard; Protein; 276 AA.

AC AAB36189;

DT 02-MAR-2001 (first entry)

DE Ehrlichia chaffeensis partial VSA5.

XX Ehrlichia chaffeensis; VSA5; variable surface antigen 5; MAP1;

XX major antigenic protein 1; antirickettsial; vaccine; gene therapy;

XX Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;

XX 3gdorf3.

XX Ehrlichia chaffeensis.

XX WO2000065063-A2.

XX 02-NOV-2000.

XX 21-APR-2000; 2000WO-US10886.

XX 22-APR-1999; 99US-0130725.

XX (UYFL) UNIV FLORIDA.

XX Barbet AF, Bowie MV, Ganta RR, Burrig MJ, Mahan SM, McGuire TC;

XX Rurangirwa FR, Moreland AL, Simbl BH, Whitmire WW, Alileman AR;

XX WPI: 2000-679675/66.

XX N-PSDB; AAC68706.

XX New polynucleotides useful as DNA vaccines for conferring immunity to

XX rickettsial infection e.g. heartwater caused by Cowdria ruminantium,

XX comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

XX Claim 3; Page 47; 63pp; English.

XX The present sequence shows a high degree of similarity to the major

XX antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be

XX used in a vaccine to protect animals or humans against rickettsial

XX diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,

CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid vaccines
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminantium genes designated map 2, lhwrfl3, 4hwrfl1, 18hwrfl1
 CC and 3gdrfl3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.

XX
 SQ Sequence 276 AA;

Query Match 85.4%; Score 88; DB 21; Length 276;
 Best Local Similarity 84.2%; Pred. No. 1.7e-06;
 Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGSGTIS 19
 ||| |||||:|||| ||
 Db 59 nttvgvfglkqndgsais 77

RESULT 9
 AAU04199
 ID AAU04199 standard; Protein: 276 AA.
 AC AAU04199;
 DT 23-OCT-2001 (first entry)
 DE Variable surface antigen 5 (VSA5) from Ehrlichia chaffeensis.
 KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
 KW infection; heartwater; diagnostic; variable surface antigen; VSA.
 XX Ehrlichia chaffeensis.
 OS US6251872-B1.
 PN 26-JUN-2001.
 PD 17-OCT-1997; 97US-09533326.
 PF 17-OCT-1996; 96US-0733230.
 PR (UYFL) UNIV FLORIDA.
 PA Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
 PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
 DR WPI; 2001-424487/45.
 DR N-PSDB; AAS07578.
 XX New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures -
 XX Example 3; Fig 2A-2B; 30pp; English.

XX The sequence represents the amino acid sequence of variable surface
 CC antigen 5 (VSA5) isolated from Ehrlichia chaffeensis, which
 CC has similarity to major antigen protein (MAP). The MAP polynucleotides
 CC and polypeptides are useful as vaccines for conferring immunity to
 CC rickettsia infection, including Cowdria ruminantium causing heartwater.
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid
 CC analysis procedures, and to produce the MAP polypeptides, which may
 CC be used to raise antibodies that are reactive with the polypeptides.
 CC The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious
 CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is
 CC needed.

SQ Sequence 276 AA;

Query Match 85.4%; Score 88; DB 22; Length 276;
 Best Local Similarity 84.2%; Pred. No. 1.7e-06;
 Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGSGTIS 19
 ||| |||||:|||| ||
 Db 59 nttvgvfglkqndgsais 77

RESULT 10
 AAW51089
 ID AAW51089 standard; Protein: 280 AA.
 AC AAW51089;
 DT 14-SEP-1998 (first entry)
 DE Ehrlichia chaffeensis major antigenic protein 1 (MAP1).
 KW MAP1 gene; major antigenic protein 1; rickettsia; DNA vaccine.
 XX Ehrlichia chaffeensis.
 OS WO9816554-A1.
 PN 23-APR-1998.
 PD 17-OCT-1997; 97WO-US19044.
 PF 17-OCT-1996; 96US-0733230.
 PR (UYFL) UNIV FLORIDA.
 PA Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 DR WPI; 1998-251232/22.
 DR N-PSDB; AAV07177.
 XX Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX Claim 3; Page 18-19; 39pp; English.

XX This polypeptide comprises the major antigen protein 1 gene (MAP1)
 CC of Ehrlichia chaffeensis. It is encoded by the MAP1 gene (see
 CC AAV07177). A claimed composition comprises a nucleic acid (see
 CC AAV07176-82) encoding a polypeptide (see AAW51088-99) that elicits a
 CC protective immune response against a rickettsial pathogen. The
 CC nucleic acid is used, in human or veterinary medicine, in vaccines
 CC to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria
 CC species. The nucleic acid does not replicate in the host but
 CC remains episomal and capable of expressing polypeptide for at least
 CC 19 mth. The Ehrlichia antigenic polypeptides can be used
 CC diagnostically to detect antibodies associated with Ehrlichia
 CC infection (claimed).

XX
 SQ Sequence 280 AA;

Query Match 85.4%; Score 88; DB 19; Length 280;
 Best Local Similarity 84.2%; Pred. No. 1.7e-06;
 Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGSGTIS 19
 ||| |||||:|||| ||
 Db 60 nttvgvfglkqndgsais 78

```

RESULT 11
AAB36183
ID AAB36183 standard; Protein; 280 AA.
XX
AC AAB36183;
XX
DT 02-MAR-2001 (first entry)
XX
DE Ehrlichia chaffeensis MAP1.
XX
KW Ehrlichia chaffeensis; MAP1; major antigenic protein 1; antirickettsial;
KW vaccine; gene therapy; Rickettsia; Cowdria; Anaplasma; map2; lhworf3;
KW 4hworf1; 18hworf1; 3gdorf3.
XX
OS Ehrlichia chaffeensis.
XX
PN WO200065063-A2.
XX
PD 02-NOV-2000.
XX
PF 21-APR-2000; 2000WO-US10886.
XX
PR 22-APR-1999; 99US-0130725.
XX
PA (UYFL ) UNIV FLORIDA.
XX
PI Barbet AF, Bowie MV, Ganta RR, Burrigide MJ, Mahan SM, McGuire TC;
PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
XX
WPI; 2000-679675/66.
DR N-PSDB; AAC68700.
XX
XX New polynucleotides useful as DNA vaccines for conferring immunity to
XX rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
XX comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
XX
PS Claim 3; Page 35-36; 63pp; English.
XX
XX The present sequence is given in a specification relating to nucleic
XX acid vaccines containing genes to protect animals or humans against
XX rickettsial diseases caused by a organisms of Rickettsia sp., Ehrlichia
XX sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
XX protective against the rickettsial pathogen. The vaccine comprises the
XX major antigenic protein 1 (MAP1) gene or the major antigenic protein 2
XX (MAP2) gene of rickettsial pathogens. The nucleic acid vaccines can be
XX ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1 and
XX 3gdorf3 may be used in therapeutic and diagnostic applications. The
XX polypeptides are useful for detecting antibodies associated with
XX infection by a rickettsial pathogen whilst the polynucleotides may be
XX used to detect the presence of rickettsial nucleic acids.
XX
SQ Sequence 280 AA;

Query Match 85.4%; Score 88; DB 21; Length 280;
Best Local Similarity 84.2%; Pred. No. 1.7e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWGSTIS 19
||| |||||:|||||
Db 60 nttvgvfglkqndgsais 78

RESULT 12
AAU04193
ID AAU04193 standard; Protein; 280 AA.
XX
AC AAU04193;
XX
DT 23-OCT-2001 (first entry)
XX
XX

```

```

DE Major antigenic protein 1 (MAP1) from Ehrlichia chaffeensis.
XX
KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
KW infection; heartwater; diagnostic.
XX
OS Ehrlichia chaffeensis.
XX
PN US6251872-B1.
XX
PD 26-JUN-2001.
XX
PF 17-OCT-1997; 97US-0953326.
XX
PR 17-OCT-1996; 96US-0733230.
XX
PA (UYFL ) UNIV FLORIDA.
XX
PI Barbet AF, Ganta RR, McGuire TC, Burrigide MJ, Nyika A;
PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX
WPI; 2001-424487/45.
DR N-PSDB; AAS07576.
XX
XX New MAP2 genes and polypeptides useful as vaccines for conferring
XX immunity to human and animal rickettsial diseases, e.g. heartwater, or
XX as molecular markers in nucleic acid analysis procedures
XX
PS Disclosure; Column 15-17; 30pp; English.
XX
XX The sequence represents the amino acid sequence of major antigenic
XX protein 1 (MAP1) from Ehrlichia chaffeensis. The MAP polynucleotides and
XX polypeptides are useful as vaccines for conferring immunity to rickettsia
XX infection, including Cowdria ruminantium causing heartwater. The MAP
XX polynucleotides may be used as molecular markers in nucleic acid
XX analysis procedures, and to produce the MAP polypeptides which may
XX be used to raise antibodies that are reactive with the polypeptides.
XX The nucleic acids may further be used as probes to identify
XX complementary sequences within other nucleic acid molecules or genomes,
XX where such probes can be applied to identify or distinguish infectious
XX strains of organisms in diagnostic procedures or in rickettsial
XX research where identification of particular organisms or strains is
XX needed.
XX
SQ Sequence 280 AA;

Query Match 85.4%; Score 88; DB 22; Length 280;
Best Local Similarity 84.2%; Pred. No. 1.7e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWGSTIS 19
||| |||||:|||||
Db 60 nttvgvfglkqndgsais 78

RESULT 13
AAU06943
ID AAU06943 standard; Protein; 281 AA.
XX
AC AAU06943;
XX
DT 05-JUL-1999 (first entry)
XX
DE E. chaffeensis OMP-1 protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
OS Ehrlichia chaffeensis.
XX
PN WO9913720-A1.
XX
PD 25-MAR-1999.

```

XX 18-SEP-1998; 98WO-US19600.
PF Ohashi N, Rikihisa Y;
PR WPI; 1999-254290/21.
XX DR N-PSDB; AAX34743.
PA Novel outer membrane proteins from Ehrlichia chaffeensis and
XX Ehrlichia canis
XX Disclosure; Fig 3B; 55pp; English.
XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the p30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 281 AA;
Query Match 85.4%; Score 88; DB 20; Length 281;
Best Local Similarity 84.2%; Pred. No. 1.7e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 NTTTGVFLKQDWGSGTIS 19
DB 59 nttvgvfglkqdwgsais 77
RESULT 14
AAY06962
ID AAY06962 standard; Protein; 280 AA.
AC AAY06962;
XX
XX 05-JUL-1999 (first entry)
XX E. canis P30-2 protein.
XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX Ehrlichia canis.
OS WO9913720-A1.
PN 25-MAR-1999.
XX 18-SEP-1998; 98WO-US19600.
PF 19-SEP-1997; 97US-0059353.
XX (OHIS) UNIV OHIO STATE.
PA Ohashi N, Rikihisa Y;
XX WPI; 1999-254290/21.
XX DR N-PSDB; AAX34762.
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX Disclosure; Fig 22B; 55pp; English.
XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part

CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the p30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 280 AA;
Query Match 77.7%; Score 80; DB 20; Length 280;
Best Local Similarity 73.7%; Pred. No. 3.5e-05;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 NTTTGVFLKQDWGSGTIS 19
DB 59 nstvgvfglkhdwnggtis 77
RESULT 15
AAY71479
ID AAY71479 standard; Protein; 280 AA.
XX
AC AAY71479;
XX
DT 12-OCT-2000 (first entry)
XX Ehrlichia canis immunoreactive protein ECA28SA3.
XX Homologous 28-kDa protein gene; ECA28SA3; immunoreactive; vaccine;
KW p28 gene; polymorphic multiple gene family; immunoprotective antigen;
KW antibacterial; canine ehrlichiosis; canine tropical pancytopenia;
KW tick-borne rickettsial disease; serodiagnosis.
XX Ehrlichia canis.
OS
FH Key Location/Qualifiers
FT Peptide 1..23
FT /label= Signal_peptide
FT Protein 24..280
FT /label= Mature_ECA28SA3_28-kDa_protein
XX WO200032745-A2.
PN 08-JUN-2000.
XX 24-NOV-1999; 99WO-US28075.
PF 30-NOV-1998; 98US-0201458.
PR 03-MAR-1999; 99US-0261358.
XX (RERE-) RES DEV FOUND.
XX Walker DH, Yu X, McBride JW;
PI WPI; 2000-412298/35.
XX N-PSDB; AAD01294, AAD01295.
XX Ehrlichia canis antigens useful for vaccinating against canine
PT ehrlichiosis in dogs -
XX
PS Claim 12; Page 68-69; 86pp; English.
XX The patent relates to homologous 28-kilodalton (kDa) protein genes of
CC Ehrlichia canis, designated ECA28SA1, ECA28SA2, ECA28SA3, ECA28-1 and
CC ECA28-2. These genes are members of a polymorphic multiple gene family
CC and contained in a single locus of 5.992 kb. The 28-kDa proteins are
CC immunoreactive with anti-E. canis serum hence are important
CC immunoprotective antigens. The protein is useful for vaccinating
CC against E. canis infections such as canine ehrlichiosis in dogs.
CC Canine ehrlichiosis, also known as canine tropical pancytopenia, is a
CC tick-borne rickettsial disease of dogs. ECA28-1 is conserved amongst
CC different strains of E. canis and hence useful for serodiagnosis of
CC canine ehrlichiosis. The present sequence is a E. canis
CC ECA28SA3 30-kDa protein which is post-translationally modified to a

CC mature 28-kDa protein by cleavage of N-terminal signal sequence.

XX
SQ Sequence 280 AA;

Query Match 77.7%; Score 80; DB 21; Length 280;
Best Local Similarity 73.7%; Pred. No. 3.5e-05;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTTGVEGLKQDWGSTIS 19
|:| ||||| ||:| |||
Db 59 nstvgvfglkhdwnggtis 77

Search completed: March 14, 2002, 09:09:38
Job time: 429 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:24:53 ; Search time 32.82 Seconds
(without alignments)
21.226 Million cell updates/sec

Title: US-09-765-739A-7

Perfect score: 103

Sequence: 1 NTTTGVFLKQDWDGSTIT 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	51	49.5	540	1 YMBM_YEAST	Q03263 saccharomyc
2	43.5	42.2	290	1 AROE_SYNY3	P74591 synchocyst
3	43	41.7	648	1 VP3_BPPH6	P11129 bacterioph
4	42.5	41.3	734	1 PURL_ZYMO	Q9req6 zymomonas m
5	41	39.8	135	1 FARE_BOVIN	P55052 bos taurus
6	41	39.8	135	1 FARE_HUMAN	Q01469 homo sapien
7	41	39.8	689	1 SYGB_PASMU	P57905 pasteurella
8	41	39.8	866	1 YCBS_ECOLI	P75857 escherichia
9	40	38.8	115	1 GCSH_CHLPN	Q9z8b0 chlamydia p
10	40	38.8	180	1 RBS_MARPA	O64416 marchantia
11	40	38.8	282	1 BIOB_HELPJ	Q9z1k8 helicobacte
12	40	38.8	282	1 BIOB_HELPY	O25956 helicobacte
13	40	38.8	377	1 YA67_METH	O27139 methanobact
14	40	38.8	942	1 ENV_CAEVG	P31627 caprine art
15	40	38.8	1025	1 DPYD_HUMAN	Q12882 homo sapien
16	40	38.8	1191	1 NKCL_SQUAC	P55013 squalus aca
17	39.5	38.3	366	1 GCST_NEIMB	Q9k018 neisseria m
18	39.5	38.3	368	1 GCST_NEIMA	Q9jvp2 neisseria m
19	39	37.9	282	1 PUNA_CELSP	P81989 cellulomona
20	39	37.9	284	1 YQAK_BACSU	P45908 bacillus su
21	39	37.9	285	1 SYGB_HAEAE	O30836 haemophilus
22	39	37.9	307	1 YMI6_MYCLE	Q32960 mycobacteri
23	39	37.9	326	1 GBLP_NICPL	P93340 nicotiana p
24	39	37.9	326	1 GBLP_TOBAC	P49026 nicotiana t
25	39	37.9	422	1 YF02_MYCPN	P75285 mycoplasma
26	39	37.9	427	1 PYRC_BACCL	P46538 bacillus ca
27	39	37.9	428	1 YB01_MYCPN	P75568 mycoplasma
28	39	37.9	438	1 YC05_MYCPN	P75571 mycoplasma
29	39	37.9	471	1 SYE_RHOSH	Q9zfa3 rhodobacter
30	39	37.9	473	1 XYLA_CLOS	P48790 clostridium
31	39	37.9	688	1 SYGB_HAEIN	P43822 haemophilus
32	39	37.9	756	1 K6PF_CAEEL	Q27483 caenorhabdi
33	39	37.9	896	1 POL_HTLIC	P14078 human t-cel

34	39	37.9	1004	1 ATNA_ARTSF	P28774 artemia san
35	39	37.9	1025	1 DPYD_BOVIN	Q28007 bos taurus
36	39	37.9	1025	1 DPYD_PIG	Q28943 sus scrofa
37	39	37.9	1203	1 YTA1_CAEEL	Q11069 caenorhabdi
38	39	37.9	1586	1 ARO1_EMENI	P07347 e pentafunc
39	38.5	37.4	536	1 FYN_XIPHE	P27446 xiphophorus
40	38	36.9	197	1 VG23_BPMD2	Q38362 mycobacteri
41	38	36.9	233	1 RR2_CYAPA	P48132 cyanophora
42	38	36.9	298	1 YMA2_MYCBO	Q02278 mycobacteri
43	38	36.9	322	1 PTNA_ECOLI	P08186 escherichia
44	38	36.9	341	1 Y33E_MYCPN	P75302 mycoplasma
45	38	36.9	371	1 Y02B_ARCFU	O30207 archaeoglob

ALIGNMENTS

RESULT 1

ID	YMBM_YEAST	STANDARD;	PRT;	540 AA.
AC	Q03263;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	HYPOTHETICAL 59.6 KDA PROTEIN IN DSK2-CAT8 INTERGENIC REGION.			
GN	YMR279C OR YMR021.05C.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972;			
RA	Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;			
RL	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.			
CC	-!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (POTENTIAL).			
CC	AS THE DRUG RESISTANCE TRANSLOCASE FAMILY). DHA12 SUBFAMILY.			
CC	STRONG, TO YEAST ATRJ.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z49704; CAA89777.1; -			
DR	SGD; S0004892; YMR279C.			
KW	Hypothetical protein; Transport; Transmembrane.			
FT	TRANSMEM 62 82 POTENTIAL.			
FT	TRANSMEM 109 129 POTENTIAL.			
FT	TRANSMEM 132 152 POTENTIAL.			
FT	TRANSMEM 170 190 POTENTIAL.			
FT	TRANSMEM 204 224 POTENTIAL.			
FT	TRANSMEM 233 253 POTENTIAL.			
FT	TRANSMEM 273 293 POTENTIAL.			
FT	TRANSMEM 296 316 POTENTIAL.			
FT	TRANSMEM 335 355 POTENTIAL.			
FT	TRANSMEM 373 393 POTENTIAL.			
FT	TRANSMEM 399 419 POTENTIAL.			
FT	TRANSMEM 430 450 POTENTIAL.			
FT	TRANSMEM 462 482 POTENTIAL.			
FT	TRANSMEM 503 523 POTENTIAL.			
SQ	SEQUENCE 540 AA; 59561 MW; 687D06CB0D70AF91 CRC64;			

Query Match 49.5%; Score 51; DB 1; Length 540;

Best Local Similarity 42.1%; Pred. No. 0.79; Indels 0; Gaps 0;

Matches 8; Conservative 3; Mismatches 8;

QY 1 NTTTGVFLKQDWDGSTIT 19

Query Match 41.3%; Score 42.5; DB 1; Length 734;
Best Local Similarity 55.6%; Pred. No. 28;
Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 3 TTGTVFLKQDW-DGSTIS 19
| | | | | : ||:
Db 551 TIGGVGLLQDWRDSTTIA 568

RESULT 5
FABE_BOVIN
ID FABE_BOVIN STANDARD; PRT; 135 AA.
AC P5052; O62808;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FATTY ACID-BINDING PROTEIN, EPIDERMAL (E-FABP) (DIFFERENTIATION-
DE ASSOCIATED LIPID BINDING PROTEIN LP2).
GN FABP5.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-33 AND 116-129.
RC TISSUE=Lens;
RX MEDLINE=97103094; PubMed=8947466;
RA Jaworski C., Wistow G.;
RT "LP2, a differentiation-associated lipid-binding protein expressed in
RT bovine lens.";
RL Biochem. J. 320:49-54(1996).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Retina;
RX MEDLINE=98198033; PubMed=9521644;
RA Kingma P.B., Bok D., Ong D.E.;
RT "Bovine epidermal fatty acid-binding protein: determination of ligand
RT specificity and cellular localization in retina and testis.";
RL Biochemistry 37:3250-3257(1998).
CC -!- FUNCTION: HIGH SPECIFICITY FOR FATTY ACIDS. HIGHEST AFFINITY FOR
CC THE UNSATURATED FATTY ACID PALMITATE (C16:0). MAY BE INVOLVED IN
CC THE UPTAKE AND TRANSPORT OF FATTY ACIDS ESSENTIAL FOR THE
CC NOURISHMENT OF THE SURROUNDING CELL TYPES. IT IS A POTENTIAL
CC TARGET FOR OXIDATIVE STRESS DURING CATARACT FORMATION IN LENS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- TISSUE SPECIFICITY: MOST ABUNDANT IN LENS AND RETINA (FOUND IN THE
CC MUELLER CELLS), MODERATELY ABUNDANT IN HEART AND TESTIS (FOUND IN
CC THE SERTOLI CELLS), AND PRESENT IN VERY LOW AMOUNTS IN LUNG.
CC -!- SIMILARITY: BELONGS TO THE FABP/P2/CBPP/CBAPP FAMILY OF
CC TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U55188; AAB41297.1; .
DR EMBL; AF059507; AAC14711.1; .
DR HSSP; P04117; 1ALB.
DR InterPro; IPR000463; Fatty_acid_BP.
DR InterPro; IPR000566; Lipocln_cytfabp.
DR Pfam; PF00061; lipocalin; 1.
DR PRINTS; PR00178; FATTYACIDBP.
DR PROSITE; PS00214; FABP; 1.
KW Transport; Lipid-binding; Phosphorylation.
FT MOD_RES 22 22 PHOSPHORYLATION (BY TYR-KINASES)
CC (PROBABLE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

FT CONFLICT 52 52 L -> P (IN REF. 1).
SQ SEQUENCE 135 AA; 15074 MW; 439B86AF88A34E2A CRC64;

Query Match 39.8%; Score 41; DB 1; Length 135;
Best Local Similarity 47.6%; Pred. No. 7.8;
Matches 10; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 NTTTGVFLKQDWG--STIS 19
| | | | | : ||:
Db 88 NFTDGLVQHEDGKKESTIT 108

RESULT 6
FABE_HUMAN
ID FABE_HUMAN STANDARD; PRT; 135 AA.
AC Q01469;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE FATTY ACID-BINDING PROTEIN, EPIDERMAL (E-FABP) (PSORIASIS-ASSOCIATED
DE FATTY ACID-BINDING PROTEIN HOMOLOG) (PA-FABP).
GN FABP5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Keratinocytes;
RX MEDLINE=92381332; PubMed=1512466;
RA Madsen P.S., Rasmussen H.H., Leffers H., Honore B., Celis J.E.;
RT "Molecular cloning and expression of a novel keratinocyte protein
RT (psoriasis-associated fatty acid-binding protein (PA-FABP)) that is
RT highly up-regulated in psoriatic skin and that shares similarity to
RT fatty acid-binding proteins.";
RL J. Invest. Dermatol. 99:299-305(1992).
RN [2]
RP SEQUENCE OF 25-33; 39-50; 62-71; 83-101 AND 120-129.
RC TISSUE=Keratinocytes;
RX MEDLINE=93162043; PubMed=1286667;
RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
RA Vandeckerckhove J.;
RT "Microsequences of 145 proteins recorded in the two-dimensional gel
RT protein database of normal human epidermal keratinocytes.";
RL Electrophoresis 13:960-969(1992).
RN [3]
RP SEQUENCE OF 67-72 AND 104-110, AND CHARACTERIZATION.
RX MEDLINE=94379963; PubMed=8092987;
RA Siegenthaler G., Hotz R., Chatellard-Grusz D., Didierjean L.,
RA Hellman U., Saurat J.-H.;
RT "Purification and characterization of the human epidermal fatty acid-
RT binding protein: localization during epidermal cell differentiation
RT in vivo and in vitro.";
RL Biochem. J. 302:363-371(1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS).
RX MEDLINE=99425110; PubMed=10493790;
RA Hohoff C., Borchers T., Rustow B., Spener F., van Tilbeurgh H.;
RT "Expression, purification and crystal structure determination of
RT recombinant human epidermal-type fatty acid-binding protein.";
RL Biochemistry 38:12229-12239(1999).
CC -!- FUNCTION: HIGH SPECIFICITY FOR FATTY ACIDS. HIGHEST AFFINITY FOR
CC C18 CHAIN LENGTH DECREASING THE CHAIN LENGTH OR INTRODUCING
CC DOUBLE BONDS REDUCES THE AFFINITY. MAY BE INVOLVED IN KERATINOCYTE
CC DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- TISSUE SPECIFICITY: KERATINOCYTES; HIGHLY EXPRESSED IN PSORIATIC
CC SKIN.
CC -!- SIMILARITY: BELONGS TO THE FABP/P2/CBPP/CBAPP FAMILY OF
CC TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

Query Match 39.8%; Score 41; DB 1; Length 866;
Best Local Similarity 63.6%; Pred. No. 59;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 8 GLKQDWGSGTI 18
| | | | | | | | | |
Db 723 GLKTDWRGTV 733

RESULT 9
GCSH_CHLPN STANDARD; PRT; 115 AA.
AC Q9Z8B0;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE GLYCINE CLEAVAGE SYSTEM H PROTEIN.
GN GCYH OR GCSH OR CPN0433 OR CP0320.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linner M., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Winn M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYSES THE DEGRADATION OF GLYCINE.
CC FROM THE P PROTEIN TO THE T PROTEIN (BY SIMILARITY).
CC -!- COFACTOR: THE H CHAIN CONTAINS A COVALENTLY-BOUND LIPOYL COFACTOR (BY SIMILARITY).
CC -!- SUBUNIT: THE GLYCINE CLEAVAGE SYSTEM IS COMPOSED OF FOUR PROTEINS: P, T, L, AND H (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE GCYH FAMILY.
CC -!- SIMILARITY: CONTAINS 1 LIPOYL-BINDING DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001626; AAD18577.1; -
CC EMBL; AE002194; AAF38175.1; -
CC EMBL; AP002546; BAA98641.1; -
CC HSSP; F16048; IHTP.
CC TIGR; CP0320; -

DR InterPro: IPR002930; GCV_H.
DR InterPro: IPR003016; Lipoyl.
DR Pfam: PF01597; GCV_H; 1.
DR PROSITE: PS00189; LIPOYL; FALSE_NEG.
KW Lipoyl; Complete proteome.
FT BINDING 58 58 LIPOYL (BY SIMILARITY).
SQ SEQUENCE 115 AA; 12876 MW; 43FF43363A48D175 CRC64;

Query Match 38.8%; Score 40; DB 1; Length 115;
Best Local Similarity 57.1%; Pred. No. 9.6;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 6 VFGLKQDWGSGTIS 19
| | | | | | | | | |
Db 97 VVRLDQDWPSNLS 110

RESULT 10
RBS_MARPA STANDARD; PRT; 180 AA.
ID RBS_MARPA
AC O64416;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN, CHLOROPLAST PRECURSOR (EC 4.1.1.39) (RUBISCO SMALL SUBUNIT).
GN RBCS.
OS Marchantia paleacea.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta; Marchantiales; Marchantiaceae; Marchantia.
OX NCBI_TaxID=56867;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VAR. DIPTERA; TISSUE=Callus;
RX MEDLINE=99205705; PubMed=10189707;
RA Suzuki T., Takio S., Tanaka K., Yamamoto I., Satoh T.;
RT "Differential light regulation of the rbcS gene expression in two cell lines of the liverwort Marchantia paleacea var. diptera.";
RL Plant Cell Physiol. 40:100-103(1999).
CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME ACTIVE SITE.
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) = 2 3-PHOSPHO-D-GLYCERATE.
CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) = 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
CC -!- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST.
CC -!- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB004883; BAA28610.1; -
DR InterPro: IPR000894; Rubisco_small.
DR Pfam: PF00101; Rubisco_small; 1.
DR ProDom; PD000290; Rubisco_small; 1.
KW Photosynthesis; Carbon dioxide fixation; Photorespiration; Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide.
FT TRANSIT 1 58 CHLOROPLAST (BY SIMILARITY).
FT CHAIN 59 180 RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN.
SQ SEQUENCE 180 AA; 19869 MW; D736A467732DCB4 CRC64;

```

Query Match      38.8%; Score 40; DB 1; Length 180;
Best Local Similarity 42.9%; Pred. No. 16;
Matches 9; Conservative 3; Mismatches 3; Indels 6; Gaps 1;

QY 5 GVFGKQDW-----DGSTIS 19
DB 37 GVFGKSEWOTKOTNGSRVS 57

RESULT 11
ID BIOB_HELPY
AC Q92JK8 STANDARD; PRT; 282 AA.
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE BACTERIA; Proteobacteria; epsilon subdivision; Helicobacter group;
GN BIOB OR JHPI298 (EC 2.8.1.6) (BIOTIN SYNTHETASE).
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923692;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Truett T.J.;
RT Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.
RL Nature 397:176-180(1999).
CC -!- CATALYTIC ACTIVITY: DETHIOBIOTIN + (S) - BIOTIN.
CC -!- PATHWAY: LAST STEP IN BIOTIN BIOSYNTHESIS PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF001553; AF006876.1;
CC InterPro: IPR002684; Biotin_synth.
CC Pfam: PF01792; Biotin_synth; 1.
CC Biotin biosynthesis; Iron-sulfur; Transferase; Complete proteome.
CC METAL 17 17 IRON-SULFUR (POTENTIAL).
CC METAL 21 21 IRON-SULFUR (POTENTIAL).
CC METAL 24 24 IRON-SULFUR (POTENTIAL).
CC SEQUENCE 282 AA; 31490 MW; 3A07177B65AFAF1B CRC64;

Query Match      38.8%; Score 40; DB 1; Length 282;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 GVFGKQDW 14
DB 157 GIFGLNESWE 166

RESULT 12
ID BIOB_HELPY
AC Q25956 STANDARD; PRT; 282 AA.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE BACTERIA; Proteobacteria; epsilon subdivision; Helicobacter group;
GN BIOB OR JHPI298 (EC 2.8.1.6) (BIOTIN SYNTHETASE).
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923692;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Truett T.J.;
RT Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.
RL Nature 397:176-180(1999).
CC -!- CATALYTIC ACTIVITY: DETHIOBIOTIN + (S) - BIOTIN.
CC -!- PATHWAY: LAST STEP IN BIOTIN BIOSYNTHESIS PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF001553; AF006876.1;
CC InterPro: IPR002684; Biotin_synth.
CC Pfam: PF01792; Biotin_synth; 1.
CC Biotin biosynthesis; Iron-sulfur; Transferase; Complete proteome.
CC METAL 17 17 IRON-SULFUR (POTENTIAL).
CC METAL 21 21 IRON-SULFUR (POTENTIAL).
CC METAL 24 24 IRON-SULFUR (POTENTIAL).
CC SEQUENCE 282 AA; 31490 MW; 3A07177B65AFAF1B CRC64;

Query Match      38.8%; Score 40; DB 1; Length 282;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 GVFGKQDW 14
DB 157 GIFGLNESWE 166

RESULT 13
ID BIOB_HELPY
AC Q27139 STANDARD; PRT; 377 AA.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE BACTERIA; Proteobacteria; epsilon subdivision; Helicobacter group;
GN BIOB OR JHPI298 (EC 2.8.1.6) (BIOTIN SYNTHETASE).
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923692;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Truett T.J.;
RT Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.
RL Nature 397:176-180(1999).
CC -!- CATALYTIC ACTIVITY: DETHIOBIOTIN + (S) - BIOTIN.
CC -!- PATHWAY: LAST STEP IN BIOTIN BIOSYNTHESIS PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF001553; AF006876.1;
CC InterPro: IPR002684; Biotin_synth.
CC Pfam: PF01792; Biotin_synth; 1.
CC Biotin biosynthesis; Iron-sulfur; Transferase; Complete proteome.
CC METAL 17 17 IRON-SULFUR (POTENTIAL).
CC METAL 21 21 IRON-SULFUR (POTENTIAL).
CC METAL 24 24 IRON-SULFUR (POTENTIAL).
CC SEQUENCE 282 AA; 31490 MW; 3A07177B65AFAF1B CRC64;

Query Match      38.8%; Score 40; DB 1; Length 282;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 GVFGKQDW 14
DB 157 GIFGLNESWE 166

RESULT 14
ID BIOB_HELPY
AC Q27139 STANDARD; PRT; 377 AA.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE BACTERIA; Proteobacteria; epsilon subdivision; Helicobacter group;
GN BIOB OR JHPI298 (EC 2.8.1.6) (BIOTIN SYNTHETASE).
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923692;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Truett T.J.;
RT Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.
RL Nature 397:176-180(1999).
CC -!- CATALYTIC ACTIVITY: DETHIOBIOTIN + (S) - BIOTIN.
CC -!- PATHWAY: LAST STEP IN BIOTIN BIOSYNTHESIS PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF001553; AF006876.1;
CC InterPro: IPR002684; Biotin_synth.
CC Pfam: PF01792; Biotin_synth; 1.
CC Biotin biosynthesis; Iron-sulfur; Transferase; Complete proteome.
CC METAL 17 17 IRON-SULFUR (POTENTIAL).
CC METAL 21 21 IRON-SULFUR (POTENTIAL).
CC METAL 24 24 IRON-SULFUR (POTENTIAL).
CC SEQUENCE 282 AA; 31490 MW; 3A07177B65AFAF1B CRC64;

Query Match      38.8%; Score 40; DB 1; Length 282;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 GVFGKQDW 14
DB 157 GIFGLNESWE 166

RESULT 15
ID BIOB_HELPY
AC Q27139 STANDARD; PRT; 377 AA.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE BACTERIA; Proteobacteria; epsilon subdivision; Helicobacter group;
GN BIOB OR JHPI298 (EC 2.8.1.6) (BIOTIN SYNTHETASE).
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923692;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Truett T.J.;
RT Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.
RL Nature 397:176-180(1999).
CC -!- CATALYTIC ACTIVITY: DETHIOBIOTIN + (S) - BIOTIN.
CC -!- PATHWAY: LAST STEP IN BIOTIN BIOSYNTHESIS PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an
```

RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.,
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- SIMILARITY: STRONG, TO M.JANNASCHII MJ1678 AND A.FULGIDUS AF0028
AND AF0181.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000877; AAB85556.1; .
DR InterPro: IPR000531; TonB_boxC.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 377 AA; 42021 MW; 24C0EE1FA77C7AE4 CRC64;

Query Match 38.8%; Score 40; DB 1; Length 377;
Best Local Similarity 35.7%; Pred. No. 35;
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 5 GVFGKQMDWGSTI 18
|||:|:|:|:
DB 223 GVLGKMKSEWEDTVL 236

RESULT 14
ENV_CAEVG STANDARD; PRT; 942 AA.
AC P31627;
DT 01-JUL-1993 (Rel. 26, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ENV POLYPROTEIN PRECURSOR (COAT POLYPROTEIN) [CONTAINS: SURFACE
DE PROTEIN; TRANSMEMBRANE PROTEIN].
GN ENV.
OS Caprine arthritis encephalitis virus (strain G63) (CAEV).
OC Viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11662;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 81-95.
RX MEDLINE=92015464; PubMed=1656067;
RA Knowles D.P. Jr., Cheevers W.P., McGuire T.C., Brassfield A.L.,
RA Harwood W.G., Stem T.A.;
RT "Structure and genetic variability of envelope glycoproteins of two
RT antigenic variants of caprine arthritis-encephalitis lentivirus.";
RL J. Virol. 65:5744-5750(1991).
RN [2]
RP REVISIONS.
RA Knowles D.P.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60855; AAB88709.2; .
DR PIR; A11307; VCLJC6.
DR InterPro: IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.

KW Glycoprotein; Coat protein; Polyprotein; Transmembrane.
FT PEPTIDE 1 80 LEADER PEPTIDE.
FT CHAIN 81 630 SURFACE PROTEIN (POTENTIAL).
FT CHAIN 631 942 TRANSMEMBRANE PROTEIN (POTENTIAL).
FT DOMAIN 1 630 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 631 659 POTENTIAL.
FT DOMAIN 660 799 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 800 820 POTENTIAL.
FT DOMAIN 821 942 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 51 91 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 131 131 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 228 228 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 348 348 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 354 354 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 379 379 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 435 435 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 441 441 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 467 467 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 481 481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 493 493 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 503 503 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 527 527 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 942 AA; 108437 MW; 5B0E1F0F3D355F4A CRC64;

Query Match 38.8%; Score 40; DB 1; Length 942;
Best Local Similarity 46.2%; Pred. No. 95;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTGTGVLGKQDM 13
|||:|:|:|:
DB 482 NTITGIMGTNTNW 494

RESULT 15
DPYD_HUMAN STANDARD; PRT; 1025 AA.
ID DPYD_HUMAN Q12882; Q16694; Q16761;
AC Q12882; Q16694; Q16761;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DIHYDROXYRIMIDINE DEHYDROGENASE [NADP+] PRECURSOR (EC 1.3.1.2) (DPD)
DE (DHPDASE) (DIHYDROURACIL DEHYDROGENASE) (DIHYDROTHYMINE
DE DEHYDROGENASE).
GN DPYD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94365020; PubMed=8083224;
RA Yokota H., Fernandez-Salguero P., Furuya H., Lin K., McBride O.W.,
RA Podschun B., Schnackerz K.D., Gonzalez F.J.;
RT "cDNA cloning and chromosome mapping of human dihydropyrimidine
RT dehydrogenase, an enzyme associated with 5-fluorouracil toxicity and
RT congenital thymine uracilluria.";
RL J. Biol. Chem. 269:23192-23196(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97280676; PubMed=9135003;
RA Johnson M.R., Wang K., Tillmanns S., Albin N., Diasio R.B.;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:10:46 ; Search time 56.96 Seconds
(without alignments)
25.409 Million cell updates/sec

Title: US-09-765-739a-7
Perfect score: 103
Sequence: 1 NTTTGVFGLKQDWGSGTIS 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	97	94.2	280	2 JE0217	28k surface antige
2	88	85.4	276	2 JE0218	28k surface antige
3	74	71.8	285	2 JE0219	28k surface antige
4	60	58.3	278	2 JE0216	28k surface antige
5	59	57.3	284	2 I40882	major antigenic pr
6	55	53.4	133	2 JE0221	28k surface antige
7	51	49.5	540	2 S54586	probable membrane
8	49	47.6	5188	2 B85547	probable RTX famil
9	47	45.6	584	2 C48658	flagellin - Escher
10	44	42.7	534	2 C82096	aminoacyl-histidin
11	44	42.7	1649	2 C86822	hypothetical prote
12	43.5	42.2	267	2 C83242	conserved hypothet
13	43.5	42.2	290	2 S76787	hypothetical prote
14	43	41.7	160	2 A75466	2-demethylmenaquin
15	43	41.7	185	2 S63457	hypothetical prote
16	43	41.7	490	2 C85354	hypothetical prote
17	43	41.7	648	1 F3BPF6	p3 protein - phage
18	42	40.8	384	2 T40502	hypothetical prote
19	41.5	40.3	293	2 A84110	sugar ABC transpor
20	41	39.8	118	2 S34346	hypothetical prote
21	41	39.8	135	2 I56326	fatty acid binding
22	41	39.8	185	2 F75255	hypothetical prote
23	41	39.8	244	2 B84885	hypothetical prote
24	41	39.8	335	2 B82220	partial fibrial u
25	41	39.8	653	2 F85620	probable outer mem
26	41	39.8	866	1 C64834	protein FLE22.12 I
27	40.5	39.3	1055	2 A96682	probable DNA (cyto
28	40.5	39.3	1559	2 T07757	hypothetical prote
29	40	38.8	85	2 T27307	hypothetical prote

30	40	38.8	115	2	G86544	glycine cleavage s
31	40	38.8	115	2	C72079	glycine cleavage s
32	40	38.8	282	2	F64695	biotin synthetase
33	40	38.8	282	2	H71823	biotin synthetase
34	40	38.8	299	2	T23932	hypothetical prote
35	40	38.8	319	2	T36845	hypothetical prote
36	40	38.8	377	2	F69008	probable membrane
37	40	38.8	404	2	S77505	acetyltransferase
38	40	38.8	408	2	S76830	probable phosphoe
39	40	38.8	470	2	JC4098	hypothetical prote
40	40	38.8	482	2	G75483	tetracycline 6-hyd
41	40	38.8	585	2	F85809	probable leucyl am
42	40	38.8	719	2	A83800	hypothetical prote
43	40	38.8	964	1	VCLJC6	penicillin-binding
44	40	38.8	1025	2	A54718	env polyprotein pr
45	40	38.8	1118	1	A49724	dihydropyrimidine
						protein-tyrosine-p

ALIGNMENTS

RESULT 1
JE0217
28k surface antigen 4 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrigge, M.J.; Alleman, A.R
C:Accession: JE0217
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0217
A:Molecule type: DNA
A:Residues: 1-280 <RED>
A:Cross-references: GB:AF062761

Query Match 94.2%; Score 97; DB 2; Length 280;
Best Local Similarity 94.7%; Pred. No. 5.5e-08;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NTTTGVFGLKQDWGSGTIS 19
Db 60 NTTTGVFGLKQDWGSGTIS 78
RESULT 2
JE0218
28k surface antigen 5 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrigge, M.J.; Alleman, A.R
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0218
A:Molecule type: DNA
A:Residues: 1-276 <RED>
A:Cross-references: GB:AF062761

Query Match 85.4%; Score 88; DB 2; Length 276;
Best Local Similarity 84.2%; Pred. No. 1.5e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NTTTGVFGLKQDWGSGTIS 19
Db 59 NTTTGVFGLKQDWGSGAIS 77


```
RESULT 8
B85547
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B85547
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85547
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GB:AE005174; NID:q12513368; PIDN:AAAG54838.1; GSPDB:GN00145; UWGP:206
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0615

Query Match 47.6%; Score 49; DB 2; Length 5188;
Best Local Similarity 53.3%; Pred. No. 60;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 TTGTGVLKQDWDGS 16
|||||:||||
Db 4803 TTSGVAAMDYDWDGA 4817

RESULT 9
C48658
flagellin - Escherichia coli (strain U5-41)
C:Species: Escherichia coli
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 26-Aug-1999
C:Accession: C48658
R:Schoenhalz, G.; Whitfield, C.
J. Bacteriol. 175, 5395-5402, 1993
A:Title: Comparative analysis of flagellin sequences from Escherichia coli strains posse
A:Reference number: A48658; MUID:93374833
A:Accession: C48658
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-584 <SCH>
A:Cross-references: GB:I07388; NID:q290438; PIDN:AAA23798.1; PID:q290439
C:Superfamily: flagellin

Query Match 45.6%; Score 47; DB 2; Length 584;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQD 12
|||||:||||
Db 199 NTTGLYGLKTE 210

RESULT 10
C82096
aminoacyl-histidine dipeptidase VC2279 [imported] - Vibrio cholerae (strain N16961 serog
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: C82096
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, B.
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: C82096
A:Status: preliminary
A:Molecule type: DNA
```

```
A:Residues: 1-534 <HEI>
A:Cross-references: GB:AE004299; GB:AE003852; NID:g9656835; PIDN:AAF95423.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2279
A:Map, position: 1
```

```
Query Match 42.7%; Score 44; DB 2; Length 534;
Best Local Similarity 70.0%; Pred. No. 32;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 TGVFGLKQDW 13
|||||:|
Db 198 TGAFLKEGW 207
```

```
RESULT 11
C86822
hypothetical protein yqbk [imported] - Lactococcus lactis subsp. lactis (strain ILL140
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
C:Accession: C86822
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarre, K.; Weissenbach, J.; Eh
Genome Res. in press, 2001
A:Title: The complete genome sequence of the lactic acid bacterium.
A:Reference number: A86625
A:Accession: C86822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1649 <STO>
A:Cross-references: GB:AE005176; NID:g12724583; PIDN:AAK05677.1; GSPDB:GN00146
A:Experimental source: strain ILL403
C:Genetics:
A:Gene: yqbk

Query Match 42.7%; Score 44; DB 2; Length 1649;
Best Local Similarity 46.7%; Pred. No. 1.1e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TTGVFGLKQDWDGST 17
|||||:||||
Db 1603 TTGTFLNQNWNSTS 1617
```

```
RESULT 12
C83242
conserved hypothetical protein PA3239 [imported] - Pseudomonas aeruginosa (strain PAO
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83242
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: C83242
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-267 <STO>
A:Cross-references: GB:AE004746; GB:AE004091; NID:g9949350; PIDN:AAG06627.1; GSPDB:GN
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA3239
```

```
Query Match 42.2%; Score 43.5; DB 2; Length 267;
Best Local Similarity 58.8%; Pred. No. 18;
Matches 10; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

Qy 1 NTTGVFGLKQDWDGST 17
```

Db 145 NTTFGVGLGL---WDPAT 158

RESULT 13

S76787

hypothetical protein - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

R:Kaneko, T.: S76787

O. K.; Okumura, S.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.

DNA Res 3:109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

s.

A:Reference number: S74322; MUID:97061201

A:Accession: S76787

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-230 (KAN)

A:Cross-references: EMBL:D90916; GB:AB001339; NID:g1653715; PIDN:BAAL8699.1; PID:g165378

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: shikimate dehydrogenase; shikimate dehydrogenase homology

F:64-269/Domain: shikimate dehydrogenase homology <SKD>

Query Match

Best Local Similarity 42.2%; Score 43.5; DB 2; Length 290;

Matches 11; Conservative 0; Mismatches 5; Indels 7; Gaps 1;

QY 2 TTTFGVFG-----LKQDWGSGT 17

I I I I I

Db 105 TNDVEGFLAPLLELKQDWSGRT 127

RESULT 14

A75466

2-demethylmenaquinone 2-C-methyltransferase (EC 2.1.1.-) DR0859 [similarity] - *Deinococcus*

C:Species: *Deinococcus radiodurans*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.W.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A:Reference number: A75250; MUID:20036896

A:Accession: A75466

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-160 (WHI)

A:Cross-references: GB:AE001940; GB:AE000513; NID:g6458577; PIDN:AAFI0437.1; PID:g645857

A:Experimental source: strain R1

C:Gene: DR0859

A:Map position: 1

C:Keywords: methyltransferase

Query Match

Best Local Similarity 41.7%; Score 43; DB 2; Length 160;

Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 GVFGGLKQDWGSGTI 18

I I I I I

Db 80 GVFGVNGEGVII 93

RESULT 15

S63457

hypothetical protein YPL025c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein LPB6c

C:Species: *Saccharomyces cerevisiae*
C:Date: 16-May-1996 #sequence_revision 12-Jul-1996 #text_change 05-Nov-1999
C:Accession: S63457
R:Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.
submitted to the EMBL Data Library, September 1995
A:Reference number: S63452
A:Accession: S63457
A:Molecule type: DNA
A:Residues: 1-185 (WAN)
A:Cross-references: EMBL:U36624; NID:g1276642; PID:g1039452; GSPDB:GN00016; MIPS:YPLC
A:Gene: MIPS:YPL025c
A:Map position: 16L
C:Superfamily: *Saccharomyces* hypothetical protein YPL025c

Query Match
Best Local Similarity 41.7%; Score 43; DB 2; Length 185;
Matches 10; Conservative 1; Mismatches 2; Indels 6; Gaps 1;

QY 1 NTTGVFGGLKQDWGSGTIS 19
I I I I I
Db 82 NQTRGM-----WDGSGTIS 94

Search completed: March 14, 2002, 09:10:47
Job time: 363 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:07:38 ; Search time 54.54 Seconds
(without alignments)
8.252 Million cell updates/sec

Title: US-09-765-739a-2
Perfect score: 110
Sequence: 1 NTTGVFGLKQDWDGATIKD 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	90	81.8	280	4	US-08-953-326-17
2	82	74.5	276	4	US-08-953-326-18
3	82	74.5	280	3	US-08-733-230-4
4	82	74.5	280	4	US-08-953-326-4
5	71	64.5	286	4	US-08-953-326-15
6	59.5	54.1	287	3	US-08-733-230-2
7	59.5	54.1	287	4	US-08-953-326-2
8	59	53.6	278	4	US-08-953-326-16
9	49	44.5	133	4	US-08-953-326-20
10	40.5	36.8	1025	2	US-08-304-309-2
11	40.5	36.8	1025	3	US-08-991-942-2
12	40.5	36.8	1025	4	US-09-138-103-2
13	40.5	36.8	1025	5	PCT-US95-04567-4
14	40	36.4	523	4	US-08-953-326-16
15	39	35.5	182	3	US-08-828-741B-2
16	39	35.5	1025	2	US-08-304-309-4
17	39	35.5	1025	3	US-08-991-942-4
18	39	35.5	1025	5	PCT-US95-04567-2
19	38.5	35.0	569	2	US-08-467-822-27
20	38.5	35.0	569	4	US-08-432-697-27
21	38.5	35.0	569	1	US-08-466-248-27
22	38.5	35.0	859	1	US-08-053-614-2
23	38.5	35.0	859	1	US-08-316-397B-2
24	38.5	35.0	859	2	US-09-034-306-2
25	38.5	35.0	859	4	US-09-259-437-2
26	38.5	35.0	859	5	PCT-US93-09782-2
27	38.5	35.0	1181	1	US-08-053-614-4

28	38.5	35.0	1181	1	US-08-316-397B-4	Sequence 4, Appli
29	38.5	35.0	1181	2	US-09-034-306-4	Sequence 4, Appli
30	38.5	35.0	1181	4	US-09-259-437-4	Sequence 4, Appli
31	38.5	35.0	1181	5	PCT-US93-09782-4	Sequence 4, Appli
32	38	34.5	135	1	US-08-446-600A-4	Sequence 4, Appli
33	38	34.5	338	4	US-09-413-231-6	Sequence 6, Appli
34	38	34.5	338	4	US-09-413-231-7	Sequence 7, Appli
35	38	34.5	371	3	US-08-586-165-3	Sequence 3, Appli
36	38	34.5	372	3	US-08-586-165-5	Sequence 5, Appli
37	38	34.5	548	1	US-08-333-358-2	Sequence 2, Appli
38	38	34.5	548	1	US-08-463-694-2	Sequence 2, Appli
39	38	34.5	548	1	US-08-694-501-2	Sequence 2, Appli
40	37.5	34.1	536	5	PCT-US95-05008-12	Sequence 12, Appli
41	37.5	34.1	567	1	US-08-967-513-5	Sequence 5, Appli
42	37.5	34.1	567	2	US-08-687-645B-5	Sequence 5, Appli
43	37.5	34.1	592	2	US-08-366-490-2	Sequence 2, Appli
44	37.5	34.1	592	2	US-08-366-490-6	Sequence 6, Appli
45	37.5	34.1	592	3	US-08-860-483A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-953-326-17
; Sequence 17, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-17

Query Match 81.8%; Score 90; DB 4; Length 280;
Best Local Similarity 88.9%; Pred. No. 4.5e-07;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGATI 18
DB 60 NTTGVFGLKQDWDGSTI 77

RESULT 2
US-08-953-326-18
; Sequence 18, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.

```

; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-18

Query Match      74.5%; Score 82; DB 4; Length 276;
Best Local Similarity 70.0%; Pred. No. 8.8e-06;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWGATIKD 20
Db 59 NNTTGVFGLKQNDGSAISN 78

RESULT 3
US-08-733-230-4
; Sequence 4, Application US/08/733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-4

```

```

Query Match      74.5%; Score 82; DB 3; Length 280;
Best Local Similarity 70.0%; Pred. No. 8.9e-06;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWGATIKD 20
Db 60 NNTTGVFGLKQNDGSAISN 79

RESULT 4
US-08-953-326-4
; Sequence 4, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4

Query Match      74.5%; Score 82; DB 4; Length 280;
Best Local Similarity 70.0%; Pred. No. 8.9e-06;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWGATIKD 20
Db 60 NNTTGVFGLKQNDGSAISN 79

RESULT 5
US-08-953-326-15
; Sequence 15, Application US/089533326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0

```

; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-15

Query Match 64.5%; Score 71; DB 4; Length 286;
Best Local Similarity 66.7%; Pred. No. 0.00055;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATI 18
||| ||||:|||||
Db 60 NTTGVFGIEQDWDRCVI 77

RESULT 6
US-08-733-230-2
; Sequence 2, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-2

Query Match 54.1%; Score 59.5; DB 3; Length 287;
Best Local Similarity 80.0%; Pred. No. 0.04;
Matches 12; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 1 NTTTGVFLKQDWDG 15
||| ||||:|||||
Db 62 NTQT-VFGLKQDWDG 75

RESULT 7

US-08-953-326-2
; Sequence 2, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
US-08-953-326-2

Query Match 54.1%; Score 59.5; DB 4; Length 287;
Best Local Similarity 80.0%; Pred. No. 0.04;
Matches 12; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 1 NTTTGVFLKQDWDG 15
||| ||||:|||||
Db 62 NTQT-VFGLKQDWDG 75

RESULT 8
US-08-953-326-16
; Sequence 16, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-16

Query Match 53.6%; Score 59; DB 4; Length 278;
Best Local Similarity 60.0%; Pred. No. 0.047;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDG 15

```
Db 60 NPTVALYGLKQDWEG 74
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1025 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-304-309-2
;
; Query Match 36.8%; Score 40.5; DB 2; Length 1025;
; Best Local Similarity 42.9%; Pred. No. 2.1e+02;
; Matches 9; Conservative 2; Mismatches 7; Indels 3; Gaps 1;
;
; OY 1 NNTTGVFLKQD---WDGATI 18
; Db 736 NTVSGLMGLKSDGTPWPAVGI 756
;
; RESULT 11
; US-08-991-942-2
; Sequence 2, Application US/08991942
; Patent No. 6015673
; GENERAL INFORMATION:
; APPLICANT: GONZALEZ, Frank J.
; APPLICANT: FERNANDEZ-SALGUERO, Pedro
; TITLE OF INVENTION: CLONING AND EXPRESSION OF CDNA FOR HUMAN
; TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,942
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/304,309
; FILING DATE: 09-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Timothy L.
; REGISTRATION NUMBER: 35,367
; REFERENCE/DOCKET NUMBER: 15280-210
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1025 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-991-942-2
;
; Query Match 36.8%; Score 40.5; DB 3; Length 1025;
; Best Local Similarity 42.9%; Pred. No. 2.1e+02;
; Matches 9; Conservative 2; Mismatches 7; Indels 3; Gaps 1;
;
; OY 1 NNTTGVFLKQD---WDGATI 18
; Db 736 NTVSGLMGLKSDGTPWPAVGI 756
;
; RESULT 10
; US-08-304-309-2
; Sequence 2, Application US/08304309
; Patent No. 5856454
; GENERAL INFORMATION:
; APPLICANT: GONZALEZ, Frank J.
; APPLICANT: FERNANDEZ-SALGUERO, Pedro
; TITLE OF INVENTION: CLONING AND EXPRESSION OF CDNA FOR HUMAN
; TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/304,309
; FILING DATE: 09-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Timothy L.
; REGISTRATION NUMBER: 35,367
; REFERENCE/DOCKET NUMBER: 15280-210
;
; Query Match 44.5%; Score 49; DB 4; Length 133;
; Best Local Similarity 56.2%; Pred. No. 0.82;
; Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
;
; OY 3 TTGVFLKQDWDGATI 18
; Db 62 TTVYVGLKENWAGDAI 77
;
; RESULT 9
; US-08-953-326-20
; Sequence 20, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangitwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; US-08-953-326-20
```


RESULT 12
US-09-138-103-2
; Sequence 2, Application US/09138103A
; Patent No. 6232448
; GENERAL INFORMATION:
; APPLICANT: Yoshikubo, Takashi
; APPLICANT: Hasegawa, Masami
; TITLE OF INVENTION: Immunological Materials and Methods for Detecting
; TITLE OF INVENTION: Dihydropyrimidine Dehydrogenase
; FILE REFERENCE: 09/138,103 Yoshikubo, et al.
; CURRENT APPLICATION NUMBER: US/09/138,103A
; CURRENT FILING DATE: 1998-08-21
; EARLIER APPLICATION NUMBER: 97114630.3
; EARLIER FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1025
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-138-103-2

Query Match 36.8%; Score 40.5; DB 4; Length 1025;
Best Local Similarity 42.9%; Pred. No. 2.1e+02;
Matches 9; Conservative 2; Mismatches 7; Indels 3; Gaps 1;

QY 1 NNTTGVFLKQD---WDGATI 18
|| :|: ||| | | |
Db 736 NTVSGLMGLKSDGTPWPAVGI 756

RESULT 13
PCT-US95-04567-4
; Sequence 4, Application PC/TUS9504567
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DIHYDROPYRIMIDINE DEHYDROGENASE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04567
; FILING DATE: CONCURRENTLY HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,357
; FILING DATE: 13-APR-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: UOAB025P--
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1025 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-04567-4

Query Match 36.8%; Score 40.5; DB 5; Length 1025;
Best Local Similarity 42.9%; Pred. No. 2.1e+02;
Matches 9; Conservative 2; Mismatches 7; Indels 3; Gaps 1;

QY 1 NNTTGVFLKQD---WDGATI 18
|| :|: ||| | | |
Db 736 NTVSGLMGLKSDGTPWPAVGI 756

RESULT 14
US-09-550-338-2
; Sequence 2, Application US/09550338
; Patent No. 6210951
; GENERAL INFORMATION:
; APPLICANT: KAWASAKI, Hisashi
; APPLICANT: USUDA, Yoshihiro
; APPLICANT: MIHARA, Yoshihiro
; APPLICANT: KURAHASHI, Osamu
; TITLE OF INVENTION: GMP Synthetase and Gene Coding for the Same
; FILE REFERENCE: 0010-1101-0
; CURRENT APPLICATION NUMBER: US/09/550,338
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: JP 11-114787
; PRIOR FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Brevibacterium lactofermentum (Corneybacterium glutamicum)
US-09-550-338-2

Query Match 36.4%; Score 40; DB 4; Length 523;
Best Local Similarity 41.2%; Pred. No. 1.1e+02;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 4 TGVFGLKQDWGATIKD 20
| :|: ||| | | :
Db 190 TEIAGLEQNWTAANIAE 206

RESULT 15
US-08-828-741B-2
; Sequence 2, Application US/08828741B
; Patent No. 6043069
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suess, Gabriele M.
; APPLICANT: Tarlington, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

Thu Mar 14 10:13:25 2002

us-09-765-739a-2.rai

Page 6

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,741B
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 182 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-828-741B-2

Query Match 35.5%; Score 39; DB 3; Length 182;
Best Local Similarity 30.0%; Pred. No. 49;
Matches 6; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
QY 1 NNTTGVFLKQDWGATIKD 20
Db 101 STDYGLILQINSRWGGLTLKE 120

Search completed: March 14, 2002, 09:07:39
Job time: 840 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:09:36 ; Search time 111.55 Seconds
(without alignments)
13.281 Million cell updates/sec

Title: US-09-765-739A-2

Perfect score: 110

Sequence: 1 NTTTGVFLKQDWGATIKD 20

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*

- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	110	100.0	288	20	AA06959		E. canis p30 prote
2	96	87.3	280	20	AA06948		E. chafeensis OMP-
3	90	81.8	280	19	AAW51094		Ehrlichia chaffeen
4	90	81.8	280	21	AAB36188		Ehrlichia chaffeen
5	90	81.8	280	22	AAU04198		Variable surface a
6	82	74.5	256	20	AA06942		E. chafeensis p28
7	82	74.5	276	19	AAW51095		Ehrlichia chaffeen
8	82	74.5	276	21	AAB36189		Ehrlichia chaffeen
9	82	74.5	276	22	AAU04199		Variable surface a
10	82	74.5	280	19	AAW51089		Ehrlichia chaffeen
11	82	74.5	280	21	AAB36183		Ehrlichia chaffeen

12	82	74.5	280	22	AAU04193	Major antigenic pr
13	82	74.5	281	20	AA06943	E. chafeensis OMP-
14	77	70.0	280	20	AA06962	E. canis p30-2 pro
15	77	70.0	280	21	AA06964	Ehrlichia canis im
16	76	69.1	276	20	AA06964	E. canis p30-4 pro
17	71	64.5	286	19	AAW51092	Ehrlichia chaffeen
18	71	64.5	286	20	AA06946	E. chafeensis OMP-
19	71	64.5	286	21	AAB36186	Ehrlichia chaffeen
20	71	64.5	286	22	AAU04196	Variable surface a
21	70	63.6	278	21	AA06947	Ehrlichia canis im
22	70	63.6	307	20	AA06961	E. canis p30-1 pro
23	59.5	54.1	287	19	AAW51088	Cowdria ruminantii
24	59.5	54.1	287	21	AAB36182	Major antigenic pr
25	59.5	54.1	287	22	AAU04192	Ehrlichia chaffeen
26	59	53.6	278	19	AAW51093	E. chafeensis OMP-
27	59	53.6	278	20	AA06947	Ehrlichia chaffeen
28	59	53.6	278	21	AAB36187	Variable surface a
29	59	53.6	278	22	AAU04197	E. chafeensis OMP-
30	59	53.6	280	20	AA06945	Variable surface a
31	49	44.5	132	22	AAU04201	Variable surface a
32	49	44.5	133	19	AAW51097	Ehrlichia canis VS
33	49	44.5	133	21	AAB36191	Ehrlichia canis pa
34	49	44.5	133	21	AA06947	Ehrlichia canis im
35	49	44.5	283	21	AA071478	Ehrlichia canis im
36	45	40.9	240	21	AAG20226	Arabidopsis thalia
37	45	40.9	364	21	AAG47084	Arabidopsis thalia
38	45	40.9	365	21	AAG47111	Arabidopsis thalia
39	45	40.9	374	21	AAG47083	Arabidopsis thalia
40	45	40.9	375	21	AAG47110	Arabidopsis thalia
41	45	40.9	442	21	AAG47082	Arabidopsis thalia
42	45	40.9	443	21	AAG47109	Arabidopsis thalia
43	45	40.9	806	22	AAG92136	C glutamicum prote
44	44	40.0	139	22	AAW24384	Human EST encoded
45	43	39.1	904	20	AAW21976	Senescence-associa

ALIGNMENTS

RESULT 1
AA06959
ID AA06959 standard: Protein; 288 AA.

AC AA06959;
DT 05-JUL-1999 (first entry)
XX
DE E. canis p30 protein.
KW Outer membrane protein; OMP; Ehrlichia chafeensis; E. canis; p30;
KW detection; dog.
XX Ehrlichia canis.
XX
PN WO9913720-Al.
XX
PD 25-MAR-1999.
PF 18-SEP-1998; 98WO-US19600.
XX 19-SEP-1997; 97US-0059353.
XX (OHIS)-UNIV-OHIO-STATE.

PI Ohashi N, Rikihisa Y;
DR WPL_1999-254290/21.
XX N-PSDB; AAX34759.
XX Novel outer membrane proteins from Ehrlichia chafeensis and
XX Ehrlichia canis
PS Disclosure; Fig 19B; 55pp; English.

Qy	1	NTTGVEGLKQDWDGAT	18
Db	60	ntttqvfqlkqdwqgsti	77

RESULT 4
AAB36188
ID AAB36188 standard; Protein; 280 AA.
XX

```

AC AAB36188;
XX
DT 02-MAR-2001 (first entry)
XX
DE Ehrlichia chaffeensis partial VSA4.
XX
KW Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAP1;
KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
KW 3gdorf3.
XX
OS Ehrlichia chaffeensis.
XX
PN WO2000065063-A2.
XX
PD 02-NOV-2000.
XX
PF 21-APR-2000; 200WO-US10886.
XX
PR 22-APR-1999; 99US-0130725.
XX
PA (UYFL ) UNIV FLORIDA.
XX
PI Barbet AF, Bowie MV, Ganta RR, Burrledge MJ, Mahan SM, McGuire TC;
PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
XX
DR WPI; 2000-679675/66.
DR N-PSDB; AAC68705.
XX
XX New polynucleotides useful as DNA vaccines for conferring immunity to
PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
PT .
XX
XX Claim 3; Page 45-46; 63pp; English.
PS
XX The present sequence shows a high degree of similarity to the major
CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
CC used in a vaccines to protect animals or humans against rickettsial
CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
CC protective against the rickettsial pathogen. The nucleic acid vaccines
CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
CC polypeptides are useful for detecting antibodies associated with
CC infection by a rickettsial pathogen whilst the polynucleotides may be
CC used to detect the presence of rickettsial nucleic acids.
XX
SQ Sequence 280 AA;

Query Match 81.8%; Score 90; DB 21; Length 280;
Best Local Similarity 88.9%; Pred. No. 1.5e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGATI 18
   ||| ||||| ||||| |||
Db 60 nttigvflgkqdwgsti 77

RESULT 5
AAU04198
ID AAU04198 standard; Protein; 280 AA.
XX
AC AAU04198;
XX
DT 23-OCT-2001 (first entry)
XX
DE Variable surface antigen 4 (VSA4) from Ehrlichia chaffeensis.
XX
KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
KW infection; heartwater; diagnostic; variable surface antigen; VSA.

```

```

XX Ehrlichia chaffeensis.
XX
PN US6251872-B1.
XX
PD 26-JUN-2001.
XX
PF 17-OCT-1997; 97US-0953326.
XX
PR 17-OCT-1996; 96US-0733230.
XX
PA (UYFL ) UNIV FLORIDA.
XX
PI Barbet AF, Ganta RR, McGuire TC, Burrledge MJ, Nyika A;
PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX
DR WPI; 2001-424487/45.
DR N-PSDB; AAS07578.
XX
XX New MAP2 genes and polypeptides useful as vaccines for conferring
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
PT as molecular markers in nucleic acid analysis procedures
PT .
XX
XX Example 3; Fig 2A-2B; 30pp; English.
PS
XX The sequence represents the amino acid sequence of variable surface
CC antigen 4 (VSA4) isolated from Ehrlichia chaffeensis, which
CC has similarity to major antigen protein (MAP). The MAP polynucleotides
CC and polypeptides are useful as vaccines for conferring immunity to
CC rickettsia infection, including Cowdria ruminantium causing heartwater.
CC The MAP polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures, and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish infectious
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
CC needed.
XX
SQ Sequence 280 AA;

Query Match 81.8%; Score 90; DB 22; Length 280;
Best Local Similarity 88.9%; Pred. No. 1.5e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGATI 18
   ||| ||||| ||||| |||
Db 60 nttigvflgkqdwgsti 77

RESULT 6
AAU06942
ID AAU06942 standard; Protein; 256 AA.
XX
AC AAU06942;
XX
DT 05-JUL-1999 (first entry)
XX
DE E. chaffeensis p28 protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
OS Ehrlichia chaffeensis.
XX
PN WO9913720-A1.
XX
PD 25-MAR-1999.
XX
PF 18-SEP-1998; 98WO-US19600.
XX

```

```

PR 19-SEP-1997; 97US-0059353.
XX (OHIS ) UNIV OHIO STATE.
XX Ohashi N, Rikihisa Y;
XX WPI; 1999-254290/21.
DR N-PSDB; AAX34742.
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
XX Ehrlichia canis
XX Claim 18; Fig 1; 55pp; English.
PS The invention provides isolated outer membrane proteins (OMP) from
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX of the OMP family and consist of proteins OMP-1, -1(B to 2) shown
XX in AY06943-958. The E. canis proteins form part of the P30 family and
XX consist of proteins shown in AY06959-970. The proteins and genes are
XX used to detect E. chaffeensis in patients and E. canis in dogs.
XX Sequence 256 AA:
SQ
Query Match 74.5%; Score 82; DB 20; Length 256;
Best Local Similarity 70.0%; Pred. NO. 2.6e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATIKD 20
Db 34 nttvgvfglkpndgsaisn 53
III IIIIIII:III: I:

RESULT 7
AAW51095
ID AAW51095 standard; Protein; 276 AA.
XX AC AAW51095;
XX DT 14-SEP-1998 (first entry)
XX DE Ehrlichia chaffeensis VSA5 protein (partial sequence).
XX KW MAP1 homologue; variable surface antigen; VSA5; rickettsia;
XX DNA vaccine.
XX OS Ehrlichia chaffeensis.
XX FH Key Location/Qualifiers
XX FT Peptide 1..25
XX FT /note= "putative signal peptide"
XX PN WO9816554-A1.
XX PD 23-APR-1998.
XX PF 17-OCT-1997; 97WO-US19044.
XX PR 17-OCT-1996; 96US-0733230.
XX (UYFL ) UNIV FLORIDA.
XX PI Barbet AF, Burrridge MJ, Ganta RR, Mahan SM, McGuire TC;
XX PI Nyika A, Rurangirwa FR;
XX DR WPI; 1998-251232/22.
XX DR N-PSDB; AAV07179.
XX Composition containing nucleic acid encoding rickettsial antigen -
XX useful for, e.g. stimulating protective immune response in humans or
XX animals
XX Claim 3; Fig 2B; 39pp; English.
PS

This is the near full-length variable surface antigen VSA5 protein
of Ehrlichia chaffeensis; it lacks about 5-7 C-terminal amino acid
residues. The VSA5 amino acid sequence was deduced from a partial
open reading frame (ORF5) of a genomic locus (see AAV07179) of E.
chaffeensis that was obtained on the basis of homology to the major
antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium. This
genomic locus included 5 ORFs encoding similar, but non-identical
proteins (see AAW51091-95). A claimed composition comprises a
nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99)
that elicits a protective immune response against a rickettsial
pathogen. The nucleic acid is used, in human or veterinary
medicine, in vaccines to protect against Rickettsia Ehrlichia,
Anaplasma and Cowdria species. The Ehrlichia antigenic
polypeptides can be used diagnostically to detect antibodies
associated with Ehrlichia infection (claimed).
XX Sequence 276 AA:
SQ
Query Match 74.5%; Score 82; DB 19; Length 276;
Best Local Similarity 70.0%; Pred. NO. 2.9e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATIKD 20
Db 59 nttvgvfglkpndgsaisn 78
III IIIIIII:III: I:

RESULT 8
AAB36189
ID AAB36189 standard; Protein; 276 AA.
XX AC AAB36189;
XX DT 02-MAR-2001 (first entry)
XX DE Ehrlichia chaffeensis partial VSA5.
XX KW Ehrlichia chaffeensis; VSA5; variable surface antigen 5; MAP1;
XX major antigenic protein 1; antirickettsial; vaccine; gene therapy;
XX Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
XX 3gdorf3.
XX OS Ehrlichia chaffeensis.
XX PN WO200065063-A2.
XX PD 02-NOV-2000.
XX PF 21-APR-2000; 2000WO-US10886.
XX PR 22-APR-1999; 99US-0130725.
XX (UYFL ) UNIV FLORIDA.
XX PI Barbet AF, Bowie MV, Ganta RR, Burrridge MJ, Mahan SM, McGuire TC;
XX PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmore WW, Alleman AR;
XX DR WPI; 2000-679675/66.
XX DR N-PSDB; AAC68706.
XX New polynucleotides useful as DNA vaccines for conferring immunity to
XX rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
XX comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
XX Claim 3; Page 47; 63pp; English.
XX The present sequence shows a high degree of similarity to the major
XX antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
XX used in a vaccines to protect animals or humans against rickettsial
XX diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,

```


DE Major antigenic protein 1 (MAP1) from Ehrlichia chaffeensis.

KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;

RW infection; heartwater; diagnostic.

OS Ehrlichia chaffeensis.

PN US6251872-B1.

PP 26-JUN-2001.

PX

PY

PZ

QA

QB

QC

QD

QE

QF

QG

QH

QI

QJ

QK

QL

QM

QN

QO

QP

QQ

QR

QS

QT

QU

QV

QW

QX

QY

QZ

RA

RB

RC

RD

RE

RF

RG

RH

RI

RJ

RK

RL

RM

RN

RO

RP

RQ

RR

RS

RT

RU

RV

RW

RX

RY

RZ

SA

SB

SC

SD

SE

SF

SG

SH

SI

SJ

SK

SL

SM

SN

SO

SP

SQ

SR

SS

ST

SV

SW

SX

SY

SZ

TA

TB

TC

TD

TE

TF

TF

TH

TI

TJ

TK

TL

TM

TN

TO

TP

TP

TR

TS

TT

TU

TU

TV

TV

TW

TX

TY

TZ

UA

UB

UC

UD

UE

UF

UG

UH

UI

UJ

UK

UL

UM

UN

UO

UP

UQ

UR

US

UT

UU

UV

UV

VW

VX

VY

VZ

WA

WB

WC

WD

WE

WF

WG

WH

WI

WJ

WK

WL

WM

WN

WO

WP

WQ

WR

WS

WT

WU

WV

WW

WX

WY

WZ

XA

XB

XC

XD

XE

XF

XF

XG

XH

XI

XJ

XK

XL

XM

XN

XO

XP

XQ

XR

XS

XT

XU

XV

XV

XW

XX

XY

XZ

YA

YB

YC

YD

YE

YF

YG

YH

YI

YJ

YK

YL

YM

YN

YO

YP

YQ

YR

YS

YT

YU

YV

YW

YX

YY

YZ

ZA

ZB

ZC

ZD

ZE

ZF

ZG

ZH

ZI

ZJ

ZK

ZL

ZM

ZN

ZO

ZP

ZQ

ZR

ZS

ZT

ZU

ZV

ZW

ZX

ZY

ZZ

AA

AB

AC

AD

AE

AF

AG

AH

AI

AJ

AK

AL

AM

AN

AO

AP

AQ

AR

AS

AT

AU

AV

AW

AX

AY

AZ

BA

BB

BC

BD

BE

BF

BG

BH

BI

BJ

BK

BL

BM

BN

BO

BP

BQ

BR

BS

BT

BU

BV

BW

BX

BY

BZ

CA

CB

CC

CD

CE

CF

CG

CH

CI

CJ

CK

CL

CM

CN

CO

CP

CQ

CR

CS

CT

CU

CV

CW

CX

CY

CZ

DA

DB

DC

DD

DE

DF

DG

DH

DI

DJ

DK

DL

DM

DN

DO

DP

DQ

DR

DS

DT

DU

DV

DW

DX

DY

DZ

EA

EB

EC

ED

EE

EF

EG

EH

EI

EJ

EK

EL

EM

EN

EO

EP

EQ

ER

ES

ET

EU

EV

EW

EX

EY

EZ

FA

FB

FC

FD

FE

FF

FG

FH

FI

FJ

FK

FL

FM

FN

FO

FP

FQ

FR

FS

FT

FU

FV

FW

FX

FY

FZ

GA

GB

GC

GD

GE

GF

GG

GH

GI

GJ

GK

GL

GM

GN

GO

GP

GQ

GR

GS

GT

GU

GV

GW

GX

GY

GZ

HA

HB

HC

HD

HE

HF

HG

HH

HI

HJ

HK

HL

HM

HN

HO

HP

HQ

HR

HS

HT

HU

HV

HW

HX

HY

HZ

IA

IB

IC

ID

IE

IF

IG

IH

II

IJ

IK

IL

IM

IN

IO

IP

IQ

IR

IS

IT

IU

IV

IW

IX

IY

IZ

JA

JB

JC

JD

JE

JF

JG

JH

JI

IJ

JK

KL

JM

JN

JO

JP

JQ

JR

JS

JT

JU

JV

JW

JX

JY

JZ

KA

KB

KC

KD

KE

KF

KG

KH

KI

KJ

KK

KL

KM

KN

KO

KP

KQ

KR

KS

KT

KU

KV

KW

KX

KY

KZ

LA

LB

LC

LD

LE

LF

LG

LH

LI

LJ

LK

LL

LM

LN

LO

LP

LQ

LR

LS

LT

LU

LV

LW

LX

LY

LZ

MA

MB

MC

MD

ME

MF

MG

MH

MI

MJ

MK

ML

MM

MN

MO

MP

MQ

MR

MS

MT

MU

MV

MW

MX

MY

MZ

NA

NB

NC

ND

NE

NF

NG

NH

NI

NJ

NK

NL

NM

NN

NO

NP

NQ

NR

NS

NT

NU

NV

NW

NX

NY

NZ

OA

OB

OC

OD

OE

OF

OG

OH

OI

OJ

OK

OL

OM

ON

OO

OP

OQ

OR

OS

OT

OU

OV

OW

OX

OY

OZ

PA

PB

PC

PD

PE

PF

PG

PH

PI

PJ

PK

PL

PM

PN

PO

PP

PQ

PR

PS

PT

PU

PV

PW

PX

PY

PZ

QA

QB

QC

XX PF 18-SEP-1998; 98WO-US19600.
 XX CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 XX CC in AAY06943-958. The E. canis proteins form part of the p30 family and
 PR 19-SEP-1997; 97US-0059353.
 XX CC consist of proteins shown in AAY06959-970. The proteins and genes are
 XX CC used to detect E. chaffeensis in patients and E. canis in dogs.
 PA (OHIS) UNIV OHIO STATE.
 XX PA Ohashi N, Rikihisa Y;
 XX PI WPI; 1999-254290/21.
 DR N-PSDB; AAX34743.
 XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and
 XX PT Ehrlichia canis
 XX PS Disclosure; Fig 3B; 55pp; English.
 XX CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the p30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX CC
 SQ Sequence 281 AA;
 Query Match 74.5%; Score 82; DB 20; Length 281;
 Best Local Similarity 70.0%; Pred. No. 2.9e-05;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NTTTGVFLKQDWDGATIKD 20
 DB 59 nttvgvfglkqndgsaisn 78
 RESULT 14
 AAY06962
 ID AAY06962 standard; Protein; 280 AA.
 XX AC AAY06962;
 XX DT 05-JUL-1999 (first entry)
 XX DE E. canis p30-2 protein.
 XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; p30;
 XX KW detection; dog.
 XX OS Ehrlichia canis.
 XX PN WO9913720-A1.
 XX PD 25-MAR-1999.
 XX PF 18-SEP-1998; 98WO-US19600.
 XX PR 19-SEP-1997; 97US-0059353.
 XX PA (OHIS) UNIV OHIO STATE.
 XX PI Ohashi N, Rikihisa Y;
 XX DR WPI; 1999-254290/21.
 DR N-PSDB; AAX34762.
 XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and
 XX PT Ehrlichia canis
 XX PS Disclosure; Fig 22B; 55pp; English.
 XX CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part

CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the p30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX CC
 SQ Sequence 280 AA;
 Query Match 70.0%; Score 77; DB 20; Length 280;
 Best Local Similarity 65.0%; Pred. No. 0.00018;
 Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 NTTTGVFLKQDWDGATIKD 20
 DB 59 nstvgvfglkhdwnggtisn 78
 RESULT 15
 AAY71479
 ID AAY71479 standard; Protein; 280 AA.
 XX AC AAY71479;
 XX DT 12-OCT-2000 (first entry)
 XX DE Ehrlichia canis immunoreactive protein ECA28SA3.
 XX KW Homologous 28-kDa protein gene; ECA28SA3; immunoreactive; vaccine;
 KW p28 gene; polymorphic multiple gene family; immunoprotective antigen;
 KW antibacterial; canine ehrlichiosis; canine tropical pancytopenia;
 KW tick-borne rickettsial disease; serodiagnosis.
 XX OS Ehrlichia canis.
 XX FH Key Location/Qualifiers
 FT Peptide 1..23
 FT /label= Signal_peptide
 FT Protein 24..280
 FT /label= Mature_ECA28SA3_28-kDa_protein
 XX PN WO200032745-A2.
 XX PD 08-JUN-2000.
 XX PF 24-NOV-1999; 99WO-US28075.
 XX PR 30-NOV-1998; 98US-0201458.
 XX PR 03-MAR-1999; 99US-0261358.
 XX PA (RERE-) RES DEV FOUND.
 XX PI Walker DH, Yu X, McBride JW;
 XX DR WPI; 2000-412298/35.
 XX DR N-PSDB; AAD01294, AAD01295.
 XX PT Ehrlichia canis antigens useful for vaccinating against canine
 XX PT ehrlichiosis in dogs -
 XX PS Claim 12; Page 68-69; 86pp; English.
 XX CC The patent relates to homologous 28-kilodalton (kDa) protein genes of
 CC Ehrlichia canis, designated ECA28SA1, ECA28SA2, ECA28SA3, ECA28-1 and
 CC ECA28-2. These genes are members of a polymorphic multiple gene family
 CC and contained in a single locus of 5.592 kb. The 28-kDa proteins are
 CC immunoreactive with anti-E. canis serum hence are important
 CC immunoprotective antigens. The protein is useful for vaccinating
 CC against E. canis infections such as canine ehrlichiosis in dogs.
 CC Canine ehrlichiosis, also known as canine tropical pancytopenia, is a
 CC tick-borne rickettsial disease of dogs. ECA28-1 is conserved amongst
 CC different strains of E. canis and hence useful for serodiagnosis of
 CC canine ehrlichiosis. The present sequence is a E. canis
 CC ECA28SA3 30-kDa protein which is post-translationally modified to a

Thu Mar 14 10:13:24 2002

us-09-765-739a-2.rag

Page 8

CC mature 28-kDa protein by cleavage of N-terminal signal sequence.
XX
SQ Sequence 280 AA;

Query Match 70.0%; Score 77; DB 21; Length 280;
Best Local Similarity 65.0%; Pred. No. 0.00018;
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 NTTTGVFGLKQDWGATIKD 20
|:|||||:|:|:|:
Db 59 nstvgvfglkhdwnggtisn 78

Search completed: March 14, 2002, 09:09:37
Job time: 428 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:24:08 ; Search time 103.37 Seconds
(without alignments)
28.301 Million cell updates/sec

Title: US-09-765-739A-2

Perfect score: 110

Sequence: 1 NTTGVFLKQWDGATIKD 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL17:*
 - 1: sp.archaea:
 - 2: sp.bacteria:
 - 3: sp.fungi:
 - 4: sp.human:
 - 5: sp.invertebrate:
 - 6: sp.mammal:
 - 7: sp.mhc:
 - 8: sp.organelle:
 - 9: sp.phage:
 - 10: sp.plant:
 - 11: sp.rodent:
 - 12: sp.virus:
 - 13: sp.vertebrate:
 - 14: sp.unclassified:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	110	100.0	288	2 Q9ZGJ2	Q9ZGJ2 ehrlichia c
2	96	87.3	280	2 Q52107	O52107 ehrlichia c
3	90	81.8	280	2 O85357	O85357 ehrlichia c
4	86	78.2	246	2 Q9RH35	Q9RH35 ehrlichia c
5	86	78.2	276	2 O85817	O85817 ehrlichia c
6	82	74.5	276	2 O85358	O85358 ehrlichia c
7	82	74.5	281	2 Q9AC19	Q9AC19 ehrlichia c
8	81	73.6	280	2 O85816	O85816 ehrlichia c
9	81	73.6	280	2 Q9ZGM9	Q9ZGM9 ehrlichia c
10	77	70.0	280	2 Q9F473	Q9F473 ehrlichia c
11	77	70.0	280	2 Q9ADV3	Q9ADV3 ehrlichia c
12	76	69.1	276	2 Q9F475	Q9F475 ehrlichia c
13	71	64.5	286	2 O52105	O52105 ehrlichia c
14	70	63.6	278	2 Q9R8A9	Q9R8A9 ehrlichia c
15	70	63.6	278	2 Q9R8A8	Q9R8A8 ehrlichia c
16	70	63.6	278	2 Q9R8A7	Q9R8A7 ehrlichia c
17	70	63.6	278	2 Q9R8A6	Q9R8A6 ehrlichia c
18	70	63.6	278	2 Q9R8A5	Q9R8A5 ehrlichia c
19	70	63.6	278	2 Q9R3J3	Q9R3J3 ehrlichia c

20	70	63.6	278	2 Q9F472	Q9F472 ehrlichia c
21	70	63.6	307	2 Q9ZGJ1	Q9ZGJ1 ehrlichia c
22	61.5	55.9	270	2 Q9AF98	Q9AF98 cowdria rum
23	61.5	55.9	290	2 Q46324	Q46324 cowdria rum
24	61.5	55.9	290	2 Q46332	Q46332 cowdria rum
25	60	54.5	290	2 Q9AEU3	Q9AEU3 cowdria rum
26	59.5	54.1	265	2 Q9AF99	Q9AF99 cowdria rum
27	59.5	54.1	287	2 Q46329	Q46329 cowdria rum
28	59.5	54.1	287	2 Q46331	Q46331 cowdria rum
29	59.5	54.1	287	2 Q9R425	Q9R425 cowdria rum
30	59.5	54.1	290	2 Q46330	Q46330 cowdria rum
31	59.5	54.1	290	2 Q46333	Q46333 cowdria rum
32	59	53.6	272	2 Q9AMF6	Q9AMF6 cowdria sp.
33	59	53.6	278	2 O52106	O52106 ehrlichia c
34	59	53.6	280	2 O52104	O52104 ehrlichia c
35	59	53.6	284	2 Q46327	Q46327 cowdria rum
36	59	53.6	284	2 Q9AF41	Q9AF41 cowdria rum
37	56	50.9	281	2 Q46328	Q46328 cowdria rum
38	56	50.9	281	2 Q9S6H1	Q9S6H1 cowdria rum
39	56	50.9	281	2 Q9S6H0	Q9S6H0 cowdria rum
40	52.5	47.3	271	2 Q9AFA0	Q9AFA0 cowdria rum
41	52	47.3	228	13 Q9I964	Q9I964 scyllorhinu
42	49	44.5	133	2 O85360	O85360 ehrlichia c
43	49	44.5	168	2 Q9EZL1	Q9EZL1 ehrlichia e
44	49	44.5	168	2 Q9EZL0	Q9EZL0 ehrlichia e
45	49	44.5	168	2 Q9EZK9	Q9EZK9 ehrlichia e

ALIGNMENTS

RESULT 1	Q9ZGJ2	PRELIMINARY;	PRT;	288 AA.
ID	Q9ZGJ2			
AC	Q9ZGJ2;			
DT	01-MAY-1999 (TREMBlrel. 10, Created)			
DT	01-MAY-1999 (TREMBlrel. 10, Last sequence update)			
DT	01-JUN-2001 (TREMBlrel. 17, Last annotation update)			
DE	30-KDA MAJOR OUTER MEMBRANE PROTEIN (P28-8).			
GN	P30 OR P28-8.			
OS	Ehrlichia canis.			
OC	Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;			
OC	Rickettsiaceae; Ehrlichieae; Ehrlichia.			
OX	NCBI TaxID=944;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-OKLAHOMA;			
RX	MEDLINE-98371112; PubMed-9705412;			
RA	Ohashi-N., Unver A., Zhi N., Rikihisa Y.;			
RT	"Cloning and characterization of multigenes encoding the			
RT	immunodominant 30-kilodalton major outer membrane proteins of			
RT	Ehrlichia canis and application of the recombinant protein for			
RT	serodiagnosis."			
RL	J. Clin. Microbiol. 36:2671-2680(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-JAKE;			
RX	MEDLINE-99242757; PubMed-10225842;			
RA	McBride J.W., Yu X.J., Walker D.H.;			
RT	"Molecular cloning of the gene for a conserved major immunoreactive			
RT	28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic			
RT	antigen."			
RL	Clin. Diagn. Lab. Immunol. 6:392-399(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-JAKE;			
RX	MEDLINE-20432107; PubMed-10974556;			
RA	McBride J.W., Yu X.J., Walker D.H.;			
RT	"A conserved, transcriptionally active p28 multigene locus of			
RT	Ehrlichia canis."			
RL	Gene 254:245-252(2000).			
DR	EMBL; AF078553; AAC68667.1;			
DR	EMBL; AF082744; AAG14362.1;			

```

DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 288 AA: 31590 MW: 86DCAECB88E9BF5E CRC64;

Query Match 100.0%; Score 110; DB 2; Length 288;
Best Local Similarity 100.0%; Pred. No. 1.9e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATIKD 20
   |||||
Db 60 NTTTGVFLKQDWDGATIKD 79

RESULT 2
ID 052107 PRELIMINARY; PRT; 280 AA.
AC 052107;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE OMP-1F.
GN OMP-1F.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
   are encoded by a polymorphic multigene family.";
RL Infect Immun. 66:133-139(1998).
DR EMBL: U72291; AAC02940.1
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA: 30731 MW: CCA6C34E2AF393E CRC64;

Query Match 87.3%; Score 96; DB 2; Length 280;
Best Local Similarity 94.4%; Pred. No. 3e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATIKD 18
   |||||
Db 60 NTTTGVFLKQDWDGATIKD 77

RESULT 3
ID 085357 PRELIMINARY; PRT; 280 AA.
AC 085357;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE 28 KDA MAJOR SURFACE ANTIGEN-4.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulisano C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Allenan A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
   the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL: AF062761; AAC26720.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
SQ SEQUENCE 280 AA: 30731 MW: CCA6C34E2AF393E CRC64;

DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA: 30743 MW: FBB841DAF08EE4DC CRC64;

Query Match 81.8%; Score 90; DB 2; Length 280;
Best Local Similarity 88.9%; Pred. No. 2.6e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATIKD 18
   |||||
Db 60 NTTTGVFLKQDWDGATIKD 77

RESULT 4
ID 09RH35 PRELIMINARY; PRT; 246 AA.
AC 09RH35;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR (FRAGMENT).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ST VINCENT;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
   human isolates of Ehrlichia chaffeensis.";
RL J Clin Microbiol. 37:1137-1143(1999).
DR EMBL: AF077735; AAC31548.1; -.
DR InterPro: IPR001702; Surface_Ag_msp4.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
FT NON_TER 1
SQ SEQUENCE 246 AA: 26884 MW: C9776392C5129A2F CRC64;

Query Match 78.2%; Score 86; DB 2; Length 246;
Best Local Similarity 83.3%; Pred. No. 9.6e-06;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATIKD 18
   |||||
Db 29 NTTAGVFLKQDWDGSAI 46

RESULT 5
ID 085817 PRELIMINARY; PRT; 276 AA.
AC 085817;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAPULPA;
RX Yu X.-J., Walker D.H.;
RT "Ehrlichia chaffeensis 28 kDa outer membrane protein.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF077734; AAC31547.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA: 30027 MW: 2FD3698FCF1F60BE CRC64;

```

Query Match 78.2%; Score 86; DB 2; Length 276;
Best Local Similarity 83.3%; Pred. No. 1.1e-05;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATI 18
DB 59 NTTAGVFLKQDWDGSAI 76

RESULT 6
OB5358
ID O85358 PRELIMINARY; PRT; 276 AA.
AC O85358;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 28 KDA MAJOR SURFACE ANTIGEN-5 (FRAGMENT).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burrige M.J.,
RA Allemen A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
the tribe Ehrlichieae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL; AF062761; AAC26716.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 276
SQ SEQUENCE 276 AA; 29782 MW; A9AB2A92263CA4EA CRC64;

Query Match 74.5%; Score 82; DB 2; Length 276;
Best Local Similarity 70.0%; Pred. No. 4.7e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATI 20
DB 59 NTTGVFLKQDWDGSAI 78

RESULT 7
Q9ACI9
ID Q9ACI9 PRELIMINARY; PRT; 281 AA.
AC Q9ACI9;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN P28.
GN P28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;

RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RL chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; U72291; AAK28673.1; -.
SQ SEQUENCE 281 AA; 30343 MW; A99E5F7C4459AA9A CRC64;

Query Match 74.5%; Score 82; DB 2; Length 281;
Best Local Similarity 70.0%; Pred. No. 4.8e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATI 20
DB 59 NTTGVFLKQDWDGSAI 78

RESULT 8
OB5816
ID O85816 PRELIMINARY; PRT; 280 AA.
AC O85816;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=91HEL17;
RA Yu X.-J., Walker D.H.;
RT "Ehrlichia chaffeensis 28 kDa outer membrane protein.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077732; AAC31545.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30277 MW; 91C54AC78507A63F CRC64;

Query Match 73.6%; Score 81; DB 2; Length 280;
Best Local Similarity 77.8%; Pred. No. 6.8e-05;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATI 18
DB 59 STTAGVFLKQDWDGSAI 76

RESULT 9
Q9ZGM9
ID Q9ZGM9 PRELIMINARY; PRT; 280 AA.
AC Q9ZGM9;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAX;
RA Yu X.-J., Walker D.H.;
RT "Ehrlichia chaffeensis 28 kDa outer membrane protein.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077733; AAC31546.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30304 MW; 91C54AC7851B7F2 CRC64;

RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
chaffeensis.";

RL Infect. Immun. 69:2083-2091(2001).

DR EMBL; AF082744; AAC14358.1; -

DR EMBL; AF324792; AAK31313.1; -

DR EMBL; AF078553; AAK28697.1; -

DR InterPro; IPR002566; Surface_Ag_msp4.

DR Pfam; PF01617; Surface_Ag_2; 1.

SQ SEQUENCE 276 AA; 30659 MW; CE51AB37D17AF3A4 CRC64;

Query Match 69.1%; Score 76; DB 2; Length 276;

Best Local Similarity 65.0%; Pred. No. 0.00041;

Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 NTTTGVFGLKQDWDGATIKD 20

DB 60 NTTTGVFGLKESWTGGIILD 79

RESULT 13

O52105 PRELIMINARY; PRT; 286 AA.

AC O52105;

DT 01-JUN-1998 (TRENBLrel. 06, Created)

DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)

DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)

DE OMP-1D.

GN OMP-1D.

OS Ehrlichia chaffeensis.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Ehrlichieae; Ehrlichia.

OX NCBI_TaxID=945;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ARKANSAS;

RX MEDLINE=98084465; PubMed=9423849;

RA Ohashi N., Zhi N., Zhang Y., Rikhiisa Y.;

RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis

are encoded by a polymorphic multigene family.";

RL Infect. Immun. 66:132-139(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-ARKANSAS;

RX MEDLINE=98321180; PubMed=9647746;

RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,

RA Allemen A.R.;

RT "Molecular characterization of a 28 kDa surface antigen gene family of

the tribe Ehrlichieae.";

RL Biochem. Biophys. Res. Commun. 247:636-643(1998).

DR EMBL; U72291; AAC02038.1; -

DR EMBL; AF062761; AAC26718.1; -

DR InterPro; IPR002566; Surface_Ag_msp4.

DR Pfam; PF01617; Surface_Ag_2; 1.

SQ SEQUENCE 286 AA; 31509 MW; F145A79270F386BE CRC64;

Query Match

Best Local Similarity 64.5%; Score 71; DB 2; Length 286;

Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 NTTTGVFGLKQDWDGATI 18

DB 60 NTTTGVFGLKESWTGGIILD 79

RESULT 14

O52105 PRELIMINARY; PRT; 278 AA.

AC O52105;

DT 01-JUN-1998 (TRENBLrel. 06, Created)

DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)

DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)

DE 28 KDA OUTER MEMBRANE PROTEIN (FRAGMENT).

OS Ehrlichia canis.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Ehrlichieae; Ehrlichia.

OX NCBI_TaxID=944;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-LOUISIANA;

RA McBride J.W., Yu X.J., Walker D.H.;

RT "Molecular cloning of a conserved major immunoreactive 28-kilodalton

protein gene from a polymorphic multiple gene family of Ehrlichia

canis.";

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF082745; AAC64551.1; -

DR InterPro; IPR002566; Surface_Ag_msp4.

DR Pfam; PF01617; Surface_Ag_2; 1.

FT NON_TER 278

SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match

Best Local Similarity 63.6%; Score 70; DB 2; Length 278;

Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 TTTTGVFGLKQDWDGATI 18

DB 61 STTVGVFGLKHDWDGSP1 77

RESULT 15

O52105 PRELIMINARY; PRT; 278 AA.

AC O52105;

DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)

DE 28 KDA OUTER MEMBRANE PROTEIN (FRAGMENT).

OS Ehrlichia canis.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Ehrlichieae; Ehrlichia.

OX NCBI_TaxID=944;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-OKLAHOMA;

RA McBride J.W., Yu X.J., Walker D.H.;

RT "Molecular cloning of a conserved major immunoreactive 28-kilodalton

protein gene from a polymorphic multiple gene family of Ehrlichia

canis.";

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF082746; AAC64552.1; -

DR InterPro; IPR002566; Surface_Ag_msp4.

DR Pfam; PF01617; Surface_Ag_2; 1.

FT NON_TER 278

SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match

Best Local Similarity 63.6%; Score 70; DB 2; Length 278;

Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 TTTTGVFGLKQDWDGATI 18

DB 61 STTVGVFGLKHDWDGSP1 77

Search completed: March 14, 2002, 09:24:09

Job time: 980 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:09:37 ; Search time 111.55 seconds
(without alignments)
13.281 Million cell updates/sec

Title: US-09-765-739A-3

Perfect score: 107

Sequence: 1 NTVGVFGLKQNDGSAISN 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_1101.*

- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	256	20	AA06942 E. chafeensis p28
2	107	100.0	276	19	AAW51095 Ehrlichia chaffeen
3	107	100.0	276	21	AAB36189 Ehrlichia chaffeen
4	107	100.0	276	22	AAU04199 Variable surface a
5	107	100.0	280	19	AAW51089 Ehrlichia chaffeen
6	107	100.0	280	21	AAB36183 Ehrlichia chaffeen
7	107	100.0	280	22	AAU04193 Major antigenic pr
8	107	100.0	281	20	AAW06943 E. chafeensis OMP-
9	91	85.0	280	19	AAW51094 Ehrlichia chaffeen
10	91	85.0	280	21	AAB36188 Ehrlichia chaffeen
11	91	85.0	280	22	AAU04198 Variable surface a

12	88	82.2	280	20	AA06948 E. chafeensis OMP-
13	82	76.6	288	20	AAW06959 E. canis p30 prote
14	80	74.8	280	20	AAW06962 E. canis p30-2 pro
15	80	74.8	280	21	AAW71479 Ehrlichia canis im
16	73	68.2	286	19	AAW51092 Ehrlichia chaffeen
17	73	68.2	286	20	AAW06946 E. chafeensis OMP-
18	73	68.2	286	21	AAB36186 Ehrlichia chaffeen
19	73	68.2	286	22	AAU04196 Variable surface a
20	72	67.3	278	21	AAW71477 Ehrlichia canis im
21	72	67.3	307	20	AAW06961 E. canis p30-1 pro
22	68	63.6	276	20	AAW06964 E. canis p30-4 pro
23	61.5	57.5	132	22	AAU04201 Variable surface a
24	61.5	57.5	133	19	AAW51097 Ehrlichia canis VS
25	61.5	57.5	133	21	AAB36191 Ehrlichia canis pa
26	61.5	57.5	133	21	AAW71480 Ehrlichia canis im
27	61.5	57.5	283	21	AAW71478 Ehrlichia canis im
28	60	56.1	278	19	AAW51093 Ehrlichia chaffeen
29	60	56.1	278	20	AAW06947 E. chafeensis OMP-
30	60	56.1	278	21	AAB36187 Ehrlichia chaffeen
31	60	56.1	278	22	AAU04197 Variable surface a
32	60	56.1	280	20	AAW06945 E. chafeensis OMP-
33	51	47.7	287	19	AAW51088 Cowdria ruminatiu
34	51	47.7	287	21	AAB36182 Cowdria ruminatiu
35	51	47.7	287	22	AAU04192 Major antigenic pr
36	44	41.1	1037	16	AAW75396 Flea sodium pump a
37	43.5	40.7	782	22	AAW93202 Human protein sequ
38	43	40.2	956	19	AAW49874 Bankia gouldi glyc
39	42	39.3	117	21	AAG46694 Arabidopsis thalia
40	42	39.3	336	21	AAG28238 Arabidopsis thalia
41	42	39.3	576	22	AAB74663 C. elegans high af
42	41	38.3	110	8	AAW70092 Sequence encoded b
43	41	38.3	111	8	AAW71705 Partial (2'-5') ol
44	41	38.3	111	16	AAW84187 Human (2'-5') olig
45	41	38.3	364	8	AAW70094 Sequence encoded b

ALIGNMENTS

RESULT 1
AA06942
ID AA06942 standard; Protein; 256 AA.

XX AC AA06942;

XX DT 05-JUL-1999 (first entry)

XX DE E. chafeensis p28 protein.

XX KW Outer membrane protein; OMP; Ehrlichia chafeensis; E. canis; p30;

XX KW detection; dog.

XX OS Ehrlichia chafeensis.

XX PN WO9913720-Al.

XX PD 25-MAR-1999.

XX PF 18-SEP-1998; 98WO-US19600.

XX PR 19-SEP-1997; 97US-0059353.

XX PA (THIS) UNIV OHIO STATE.

XX PI Ohashi N, Rikihisa Y;

XX DR WPI; 1999-254290/21

XX N-PSDB: AA024742.

XX PT Novel outer membrane proteins from Ehrlichia chafeensis and

XX PT Ehrlichia canis

XX PS Claim 18; Fig 1; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY05943-958. The E. canis proteins form part of the p30 family and
 CC consist of proteins shown in AAY05959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX
 SQ Sequence 256 AA;

Query Match 100.0%; Score 107; DB 20; Length 256;
 Best Local Similarity 100.0%; Pred. No. 2.5e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTVGVFLKQKQNDGSAISN 20
 |||||
 Db 34 nttvgvfglkqndgsaaisn 53

RESULT 2
 AAW51095
 ID AAW51095 standard; Protein: 276 AA.
 AC AAW51095;

DT 14-SEP-1998 (first entry)

DE Ehrlichia chaffeensis VSA5 protein (partial sequence).

KW MAP1 homologue; variable surface antigen; VSA5; rickettsia;
 KW DNA vaccine.

OS Ehrlichia chaffeensis.

XX Key Location/Qualifiers

FT Peptide 1..25 /note= "putative signal peptide"

XX W09816554-AL.

XX 23-APR-1998.

XX 17-OCT-1997; 97WO-US19044.

XX 17-OCT-1996; 96US-0733230.

XX (UYFL) UNIV FLORIDA.

PI Barbet AF, Burrigide MJ, Ganta RR, Mahan SM, McGuire TC;

PI Nyika A, Rurangirwa FR;

DR WPI: 1998-251232/22.

DR N-PSDB; AAV07179.

XX Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals

PS Claim 3; Fig 2B; 39pp; English.

XX This is the near full-length variable surface antigen VSA5 protein
 CC of Ehrlichia chaffeensis; it lacks about 5-7 C-terminal amino acid
 CC residues. The VSA5 amino acid sequence was deduced from a partial
 CC open reading frame (ORF5) of a genomic locus (see AAV07179) of E.
 CC chaffeensis that was obtained on the basis of homology to the major
 CC antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium. This
 CC genomic locus included 5 ORFs encoding similar, but non-identical
 CC proteins (see AAW51091-95). A claimed composition comprises a
 CC nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99)
 CC that elicits a protective immune response against a rickettsial
 CC pathogen. The nucleic acid is used, in human or veterinary
 CC medicine, in vaccines to protect against Rickettsia, Ehrlichia,

CC Anaplasma and Cowdria species. The Ehrlichia antigenic
 CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).

SQ Sequence 276 AA;

Query Match 100.0%; Score 107; DB 19; Length 276;
 Best Local Similarity 100.0%; Pred. No. 2.7e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTVGVFLKQKQNDGSAISN 20
 |||||
 Db 59 nttvgvfglkqndgsaaisn 78

RESULT 3

AAB36189

ID AAB36189 standard; Protein: 276 AA.

XX AAB36189;

XX 02-MAR-2001 (first entry)

XX Ehrlichia chaffeensis partial VSA5.

KW Ehrlichia chaffeensis; VSA5; variable surface antigen 5; MAP1;

KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;

KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;

KW 3gdorf3.

XX Ehrlichia chaffeensis.

XX W0200065063-A2.

XX 02-NOV-2000.

XX 21-APR-2000; 2000WO-US10886.

XX 22-APR-1999; 99US-0130725.

XX (UYFL) UNIV FLORIDA.

XX Barbet AF, Bowle MV, Ganta RR, Burrigide MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;

XX WPI: 2000-679675/66.

XX N-PSDB; AAC68706.

XX New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

PS Claim 3; Page 47; 63pp; English.

XX The present sequence shows a high degree of similarity to the major
 CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
 CC used in a vaccine to protect animals or humans against rickettsial
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid vaccines
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
 CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.

SQ Sequence 276 AA;

Query Match 100.0%; Score 107; DB 21; Length 276;

Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
|||||
Db 59 nttvgvfglkqwdgsaisn 78

RESULT 4

ID AAU04199 standard; Protein; 276 AA.

AC AAU04199;

DT 23-OCT-2001 (first entry)

DE Variable surface antigen 5 (VSA5) from Ehrlichia chaffeensis.

KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
infection; heartwater; diagnostic; variable surface antigen; VSA.

OS Ehrlichia chaffeensis.

PN US251872-B1.

PD 26-JUN-2001.

PF 17-OCT-1997; 97US-0953326.

PR 17-OCT-1996; 96US-0733230.

PA (UYFL) UNIV FLORIDA.

PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
PI Rutangirwa FR, Mahan SM, Bowie MV, Alleman AR;

DR WPI; 2001-424487/45.

DR N-PSDB; AAS07578.

PT New MAP2 genes and polypeptides useful as vaccines for conferring
immunity to human and animal rickettsial diseases, e.g. heartwater, or
as molecular markers in nucleic acid analysis procedures -

PS Example 3; Fig 2A-2B; 30pp; English.

CC The sequence represents the amino acid sequence of variable surface
antigen 5 (VSA5) isolated from Ehrlichia chaffeensis, which
has similarity to major antigen protein (MAP). The MAP polynucleotides
and polypeptides are useful as vaccines for conferring immunity to
rickettsia infection, including Cowdria ruminantium causing heartwater.
The MAP polynucleotides may be used as molecular markers in nucleic acid
analysis procedures, and to produce the MAP polypeptides, which may
be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
complementary sequences within other nucleic acid molecules or genomes,
where such probes can be applied to identify or distinguish infectious
strains of organisms in diagnostic procedures or in rickettsial
research where identification of particular organisms or strains is
needed.

SQ Sequence 276 AA;

Query Match 100.0%; Score 107; DB 22; Length 276;

Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
|||||
Db 59 nttvgvfglkqwdgsaisn 78

RESULT 5

AAW51089

ID AAW51089 standard; Protein; 280 AA.

AC AAW51089;

DT 14-SEP-1998 (first entry)

DE Ehrlichia chaffeensis major antigenic protein 1 (MAP1).

KW MAP1 gene; major antigenic protein 1; rickettsia; DNA vaccine.

OS Ehrlichia chaffeensis.

PN WO9816554-A1.

PD 23-APR-1998.

PF 17-OCT-1997; 97WO-US19044.

PR 17-OCT-1996; 96US-0733230.

PA (UYFL) UNIV FLORIDA.

PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
PI Nyika A, Rutangirwa FR;

DR WPI; 1998-251232/22.

DR N-PSDB; AAV07177.

PT Composition containing nucleic acid encoding rickettsial antigen -
useful for, e.g. stimulating protective immune response in humans or
animals

PS Claim 3; Page 18-19; 39pp; English.

CC This polypeptide comprises the major antigen protein 1 gene (MAP1)
of Ehrlichia chaffeensis. It is encoded by the MAP1 gene (see
CC AAV07177). A claimed composition comprises a nucleic acid (see
CC AAV07176-82) encoding a polypeptide (see AAW51088-99) that elicits a
CC protective immune response against a rickettsial pathogen. The
CC nucleic acid is used, in human or veterinary medicine, in vaccines
CC to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria
CC species. The nucleic acid does not replicate in the host but
CC remains episomal and capable of expressing polypeptide for at least
CC 19 mth. The Ehrlichia antigenic polypeptides can be used
CC diagnostically to detect antibodies associated with Ehrlichia
infection (claimed).

SQ Sequence 280 AA;

Query Match 100.0%; Score 107; DB 19; Length 280;

Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
|||||
Db 60 nttvgvfglkqwdgsaisn 79

RESULT 6

ID AAB36183 standard; Protein; 280 AA.

AC AAB36183;

DT 02-MAR-2001 (first entry)

DE Ehrlichia chaffeensis MAP1.

KW Ehrlichia chaffeensis; MAP1; major antigenic protein 1; antirickettsial;
vaccine; gene therapy; Rickettsia; Cowdria; Anaplasma; map2; lhworf3;
4hworfl; 18hworfl; 3gdorf3.

```

XX OS Ehrlichia chaffeensis.
XX PN WO200065063-A2.
XX PD 02-NOV-2000.
XX PF 21-APR-2000; 2000WO-US10886.
XX PR 22-APR-1999; 99US-0130725.
XX XX (UYFL ) UNIV FLORIDA.
XX PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, McGuire TC;
XX PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
XX DR N-PSDB; AAC68700.
XX XX WPI: 2000-679675/66.
XX DR N-PSDB; AAC68700.
XX XX New polynucleotides useful as DNA vaccines for conferring immunity to
XX PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
XX PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
XX PT .
XX PS Claim 3; Page 35-36; 63pp; English.
XX XX The present sequence is given in a specification relating to nucleic
XX CC acid vaccines containing genes to protect animals or humans against
XX CC rickettsial diseases caused by a organisms of Rickettsia sp., Ehrlichia
XX CC sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
XX CC protective against the rickettsial pathogen. The vaccine comprises the
XX CC major antigenic protein 1 (MAP1) gene or the major antigenic protein 2
XX CC (MAP2) gene of rickettsial pathogens. The nucleic acid vaccines can be
XX CC driven by the human cytomegalovirus (HCMV) enhancer-promoter. Cowdria
XX CC ruminantium genes designated map 2, lbworf3, 4hworf1, 18hworf1 and
XX CC 3gdoorf3 may be used in therapeutic and diagnostic applications. The
XX CC polypeptides are useful for detecting antibodies associated with
XX CC infection by a rickettsial pathogen whilst the polynucleotides may be
XX CC used to detect the presence of rickettsial nucleic acids.
XX XX Sequence 280 AA;

Query Match 100.0%; Score 107; DB 21; Length 280;
Best Local Similarity 100.0%; Pred. No. 2,7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNMDSAISN 20
Db |||||
60 nttvgvfglkqndgsaisn 79

RESULT 7
AAU04193
ID AAU04193 standard; Protein; 280 AA.
XX AC AAU04193;
XX DT 23-OCT-2001 (first entry)
XX DE Major antigenic protein 1 (MAP1) from Ehrlichia chaffeensis.
XX KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
XX KM infection; heartwater; diagnostic.
XX XX Ehrlichia chaffeensis.
XX OS US6251872-B1.
XX PN 26-JUN-2001.
XX PD 17-OCT-1997; 97US-0953326.
XX PF
XX XX

Query Match 100.0%; Score 107; DB 22; Length 280;
Best Local Similarity 100.0%; Pred. No. 2,7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNMDSAISN 20
Db |||||
60 nttvgvfglkqndgsaisn 79

RESULT 8
AAY06943
ID AAY06943 standard; Protein; 281 AA.
XX AC AAY06943;
XX DT 05-JUL-1999 (first entry)
XX DE E. chaffeensis OMP-1 protein.
XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX KM detection; dog.
XX OS Ehrlichia chaffeensis.
XX PN WO9913720-A1.
XX PD 25-MAR-1999.
XX PF 18-SEP-1998; 98WO-US19600.
XX XX 19-SEP-1997; 97US-0059353.
XX PR (OHIS ) UNIV OHIO STATE.
XX PA Ohashi N, Rikihisa Y;
XX PI WPI: 1999-254290/21.
XX DR N-PSDB; AAX34743.
XX XX Novel outer membrane proteins from Ehrlichia chaffeensis and
XX PT

```

```

PR 17-OCT-1996; 96US-0733230.
XX XX (UYFL ) UNIV FLORIDA.
XX PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
XX PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX DR WPI: 2001-424487/45.
XX DR N-PSDB; AAS07576.
XX XX New MAP2 genes and polypeptides useful as vaccines for conferring
XX PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
XX PT as molecular markers in nucleic acid analysis procedures
XX PS Disclosure; Column 15-17; 30pp; English.
XX XX The sequence represents the amino acid sequence of major antigenic
XX CC protein 1 (MAP1) from Ehrlichia chaffeensis. The MAP polynucleotides and
XX CC polypeptides are useful as vaccines for conferring immunity to rickettsia
XX CC infection, including Cowdria ruminantium causing heartwater. The MAP
XX CC polynucleotides may be used as molecular markers in nucleic acid
XX CC analysis procedures, and to produce the MAP polypeptides, which may
XX CC be used to raise antibodies that are reactive with the polypeptides.
XX CC The nucleic acids may further be used as probes to identify
XX CC complementary sequences within other nucleic acid molecules or genomes,
XX CC where such probes can be applied to identify or distinguish infectious
XX CC strains of organisms in diagnostic procedures or in rickettsial
XX CC research where identification of particular organisms or strains is
XX CC needed.
XX XX Sequence 280 AA;

Query Match 100.0%; Score 107; DB 22; Length 280;
Best Local Similarity 100.0%; Pred. No. 2,7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNMDSAISN 20
Db |||||
60 nttvgvfglkqndgsaisn 79

RESULT 8
AAY06943
ID AAY06943 standard; Protein; 281 AA.
XX AC AAY06943;
XX DT 05-JUL-1999 (first entry)
XX DE E. chaffeensis OMP-1 protein.
XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX KM detection; dog.
XX OS Ehrlichia chaffeensis.
XX PN WO9913720-A1.
XX PD 25-MAR-1999.
XX PF 18-SEP-1998; 98WO-US19600.
XX XX 19-SEP-1997; 97US-0059353.
XX PR (OHIS ) UNIV OHIO STATE.
XX PA Ohashi N, Rikihisa Y;
XX PI WPI: 1999-254290/21.
XX DR N-PSDB; AAX34743.
XX XX Novel outer membrane proteins from Ehrlichia chaffeensis and
XX PT

```

PT Ehrlichia canis
 PS Disclosure; Fig 3B; 55pp; English.
 XX
 CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the p30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX
 SQ Sequence 281 AA;
 Query Match 100.0%; Score 107; DB 20; Length 281;
 Best Local Similarity 100.0%; Pred. No. 2.8e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NTTGVGFLKQNWGSAISN 20
 Db 59 nttvgvfglkqngdgsaisn 78
 RESULT 9
 AAW51094
 ID AAW51094 standard; Protein; 280 AA.
 XX
 AC AAW51094;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Ehrlichia chaffeensis VSA4 protein.
 XX
 KW MAPI homologue; variable surface antigen; VSA4; rickettsia;
 KW DNA vaccine.
 XX
 OS Ehrlichia chaffeensis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..25
 FT /note= "putative signal peptide"
 XX
 PN W09816554-A1.
 XX
 PD 23-APR-1998.
 XX
 PF 17-OCT-1997; 97WO-US19044.
 XX
 PR 17-OCT-1996; 96US-0733230.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Burrige MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 XX
 DR WPI; 1998-251232/22.
 DR N-PSDB; AAV07179.
 XX
 PS Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX
 PS Claim 3; Fig 2B; 39pp; English.
 XX
 CC This is the full-length variable surface antigen VSA4 protein of
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
 CC partial open reading frame (ORF4) of a genomic locus (see AAV07179)
 CC of E. chaffeensis that was obtained on the basis of homology to the
 CC major antigenic protein MAPI (see AAW51088) of Cowdria ruminantium.
 CC This genomic locus included 5 ORFs encoding similar, but
 CC non-identical proteins (see AAW51091-95). A claimed composition
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
 CC (see AAW51088-99) that elicits a protective immune response against a

CC rickettsial pathogen. The nucleic acid is used, in human or
 CC veterinary medicine, in vaccines to protect against Rickettsia,
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
 CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).
 XX
 SQ Sequence 280 AA;
 Query Match 85.0%; Score 91; DB 19; Length 280;
 Best Local Similarity 84.2%; Pred. No. 1e-06;
 Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 NTTGVGFLKQNWGSAIS 19
 Db 60 nttvgvfglkqngdgsstis 78
 RESULT 10
 AAB36188
 ID AAB36188 standard; Protein; 280 AA.
 XX
 AC AAB36188;
 XX
 DT 02-MAR-2001 (first entry)
 XX
 DE Ehrlichia chaffeensis partial VSA4.
 XX
 KW Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAPI;
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
 KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
 KW 3gdorf3.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN WO2000065063-A2.
 XX
 PD 02-NOV-2000.
 XX
 PF 21-APR-2000; 2000WO-US10886.
 XX
 PR 22-APR-1999; 99US-0130725.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Bowie MV, Ganta RR, Burrige MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmore WW, Allemen AR;
 XX
 DR WPI; 2000-679675/66.
 DR N-PSDB; AAC68705.
 XX
 PT New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
 XX
 PS Claim 3; Page 45-46; 63pp; English.
 XX
 CC The present sequence shows a high degree of similarity to the major
 CC antigenic protein 1 (MAPI) of Ehrlichia sp. The MAPI gene may be
 CC used in a vaccines to protect animals or humans against rickettsial
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid vaccines
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
 CC and 3gdorf3 may be useful in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.
 XX
 SQ Sequence 280 AA;

Query Match 85.0%; Score 91; DB 21; Length 280;
 Best Local Similarity 84.2%; Pred. No. 1e-06; 1; Mismatches 0; Gaps 0;
 Matches 16; Conservative 2;

QY 1 NTTGVFGLKQNWGSAIS 19
 DB 60 nttgvgfglkqwdgstis 78
 |||:|||||:|||||

RESULT 11
 AAU04198
 ID AAU04198 standard; Protein; 280 AA.
 XX AC AAU04198;
 XX DT 23-OCT-2001 (first entry)
 XX DE Variable surface antigen 4 (VSA4) from Ehrlichia chaffeensis.
 XX KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
 XX KW infection; heartwater; diagnostic; variable surface antigen; VSA.
 XX OS Ehrlichia chaffeensis.
 XX PN US6251872-B1.
 XX PD 26-JUN-2001.
 XX PF 17-OCT-1997; 97US-0953326.
 XX PR 17-OCT-1996; 96US-0733230.
 XX PA (UYFL) UNIV FLORIDA.
 XX PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
 XX PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
 XX DR WPI; 2001-424487/45.
 XX DR N-PSDB; AAS07578.
 XX PT New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures
 XX PS Example 3; Fig 2A-2B; 30pp; English.

CC The sequence represents the amino acid sequence of variable surface
 CC antigen 4 (VSA4) isolated from Ehrlichia chaffeensis, which
 CC has similarity to major antigen protein (MAP). The MAP polynucleotides
 CC and polypeptides are useful as vaccines for conferring immunity to
 CC rickettsia infection, including Cowdria ruminantium causing heartwater.
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid
 CC analysis procedures, and to produce the MAP polypeptides, which may
 CC be used to raise antibodies that are reactive with the polypeptides.
 CC The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious
 CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is
 CC needed.

XX Sequence 280 AA;

Query Match 85.0%; Score 91; DB 22; Length 280;
 Best Local Similarity 84.2%; Pred. No. 1e-06; 1; Mismatches 0; Gaps 0;
 Matches 16; Conservative 2;

QY 1 NTTGVFGLKQNWGSAIS 19
 DB 60 nttgvgfglkqwdgstis 78
 |||:|||||:|||||

RESULT 12
 AAY06948
 ID AAY06948 standard; Protein; 280 AA.
 XX AC AAY06948;
 XX DT 05-JUL-1999 (first entry)
 XX DE E. chaffeensis OMP-1F protein.
 XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 XX KW detection; dog.
 XX OS Ehrlichia chaffeensis.
 XX PN WO9913720-A1.
 XX PD 25-MAR-1999.
 XX PF 18-SEP-1998; 98WO-US19600.
 XX PR 19-SEP-1997; 97US-0059353.
 XX PA (OHIS) UNIV OHIO STATE.
 XX PI Ohashi N, Rikihisa Y;
 XX DR WPI; 1999-254290/21.
 XX DR N-PSDB; AAX34748.
 XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis
 XX PS Claim 16; Fig 8B; 55pp; English.
 XX CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 280 AA;

Query Match 82.2%; Score 88; DB 20; Length 280;
 Best Local Similarity 84.2%; Pred. No. 3.1e-06;
 Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAIS 19
 DB 60 nttgvgfglkqwdgstis 78
 |||:|||||:|||||

RESULT 13
 AAY06959
 ID AAY06959 standard; Protein; 288 AA.
 XX AC AAY06959;
 XX DT 05-JUL-1999 (first entry)
 XX DE E. canis P30 protein.
 XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 XX KW detection; dog.
 XX OS Ehrlichia canis.
 XX PN WO9913720-A1.
 XX PD 25-MAR-1999.

XX 18-SEP-1998; 98WO-US19600.
 XX
 PR 19-SEP-1997; 97US-0059353.
 XX
 PA (OHIS) UNIV OHIO STATE.
 XX

PI Ohashi N, Rikihisa Y;
 XX WPI; 1999-254290/21.
 DR N-PSDB; AAX34759.
 XX
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis
 XX
 PS Disclosure; Fig 19B; 55pp; English.
 XX

CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the p30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX

SQ Sequence 288 AA;

Query Match 76.6%; Score 82; DB 20; Length 288;
 Best Local Similarity 70.0%; Pred. No. 3e-05;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 NTTGVFGLKQNWGSAISN 20
 IIIIIIIII:II: I:
 DB 60 nttgvgfllkqdwgdgatkld 79

RESULT 14
 AAY06962
 ID AAY06962 standard; Protein; 280 AA.
 XX
 AC AAY06962;
 XX

DT 05-JUL-1999 (first entry)

DE E. canis P30-2 protein.

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.

OS Ehrlichia canis.

PN WO9913720-A1.

PD 25-MAR-1999.

PF 18-SEP-1998; 98WO-US19600.

PR 19-SEP-1997; 97US-0059353.

PA (OHIS) UNIV OHIO STATE.

PI Ohashi N, Rikihisa Y;

DR WPI; 1999-254290/21.

DR N-PSDB; AAX34762.

XX Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis

PS Disclosure; Fig 22B; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part

CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the p30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX

SQ Sequence 280 AA;

Query Match 74.8%; Score 80; DB 20; Length 280;
 Best Local Similarity 70.0%; Pred. No. 6e-05;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 NTTGVFGLKQNWGSAISN 20
 I:IIIIIIII:II: III
 DB 59 nstvgvfgllkhdwnggtisn 78

RESULT 15
 AAY71479
 ID AAY71479 standard; Protein; 280 AA.
 XX

AC AAY71479;

DT 12-OCT-2000 (first entry)

DE Ehrlichia canis immunoreactive protein ECA28SA3.

KW Homologous 28-kDa protein gene; ECA28SA3; immunoreactive; vaccine;
 KW p28 gene; polymorphic multiple gene family; immunoprotective antigen;
 KW antibacterial; canine ehrlichiosis; canine tropical pancytopenia;
 KW tick-borne rickettsial disease; serodiagnosis.

OS Ehrlichia canis.

FH Key Location/Qualifiers

FT Peptide 1..23

FT /label= Signal_peptide

FT Protein 24..280

FT /label= Mature_ECA28SA3_28-kDa_protein

PN WO200032745-A2.

PD 08-JUN-2000.

PF 24-NOV-1999; 99WO-US28075.

PR 30-NOV-1998; 98US-0201458.

PR 03-MAR-1999; 99US-0261358.

PA (RERE-) RES DEV FOUND.

PI Walker DH, Yu X, McBride JW;

DR WPI; 2000-412298/35.

DR N-PSDB; AAD01294, AAD01295.

XX Ehrlichia canis antigens useful for vaccinating against canine
 PT ehrlichiosis in dogs -

PS Claim 12; Page 68-69; 86pp; English.

XX The patent relates to homologous 28-kilodalton (kDa) protein genes of
 CC Ehrlichia canis, designated ECA28SA1, ECA28SA2, ECA28SA3, ECA28-1 and
 CC ECA28-2. These genes are members of a polymorphic multiple gene family
 CC and contained in a single locus of 5.592 kb. The 28-kDa proteins are
 CC immunoreactive with anti-E. canis serum hence are important
 CC against E. canis infections such as canine ehrlichiosis in dogs.
 CC Canine ehrlichiosis, also known as canine tropical pancytopenia, is a
 CC tick-borne rickettsial disease of dogs. ECA28-1 is conserved amongst
 CC different strains of E. canis and hence useful for serodiagnosis of
 CC canine ehrlichiosis. The present sequence is a E. canis
 CC ECA28SA3 30-kDa protein which is post-translationally modified to a

CC mature 28-kDa protein by cleavage of N-terminal signal sequence.

XX
SQ Sequence 280 AA;

Query Match 74.8%; Score 80; DB 21; Length 280;
Best Local Similarity 70.0%; Pred. No. 6e-05;
Matches 14; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTVGVFGLKQNWGSAISN 20
|:|||||:|:|
Db 59 nstvgvfglkhdwnggtisn 78

Search completed: March 14, 2002, 09:09:37
Job time: 428 Sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:24:09 ; Search time 103.37 seconds
(without alignments)
25.471 Million cell updates/sec

Title: US-09-765-739A-4
Perfect score: 98
Sequence: 1 NPTVALYGLKQDNGVSA 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing: first 45 summaries

- Database : SPTREMBL_17:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query %		Length	DB	ID
		Match				
1	98	100.0	280	2	O52104	ehrlchia c
2	88	89.8	278	2	O52106	ehrlchia c
3	62	63.3	280	2	O9f473	ehrlchia c
4	62	63.3	280	2	O9ADV3	ehrlchia c
5	61	62.2	280	2	O85357	ehrlchia c
6	59	60.2	288	2	O9ZGJ2	ehrlchia c
7	58	59.2	246	2	O9RH35	ehrlchia c
8	58	59.2	276	2	O85817	ehrlchia c
9	58	59.2	280	2	O52107	ehrlchia c
10	57	58.2	276	2	O85358	ehrlchia c
11	57	58.2	281	2	O9ACI9	ehrlchia c
12	55	56.1	272	2	O9AMF6	ehrlchia c
13	55	56.1	284	2	O46327	ehrlchia sp.
14	55	56.1	284	2	O9AFAL	ehrlchia c
15	53	54.1	280	2	O85816	ehrlchia c
16	53	54.1	280	2	O9ZGM9	ehrlchia c
17	52	53.1	278	2	O9R8A9	ehrlchia c
18	52	53.1	278	2	O9R8A8	ehrlchia c
19	52	53.1	278	2	O9R8A7	ehrlchia c

20	52	53.1	278	2	O9R8A6	ehrlchia c	Q9r8a6 ehrlchia c Q9r8a5 ehrlchia c Q9r3j3 ehrlchia c Q9f472 ehrlchia c Q9g9j1 ehrlchia c Q9af99 cowdria rum Q9af98 cowdria rum Q46328 cowdria rum Q9s6h1 cowdria rum Q9s6h0 cowdria rum Q46329 cowdria rum Q46331 cowdria rum Q9f425 cowdria rum Q46324 cowdria rum Q46330 cowdria rum Q46332 cowdria rum Q46333 cowdria rum Q9aeu3 cowdria rum O52105 ehrlchia c O08347 streptomyce Q9f475 ehrlchia c Q9klm9 vibrio chol Q9f7n4 uncultured Q9afao cowdria rum O85360 ehrlchia c Q9f474 ehrlchia c
21	52	53.1	278	2	O9R8A5	ehrlchia c	
22	52	53.1	278	2	O9R3J3	ehrlchia c	
23	52	53.1	278	2	O9F472	ehrlchia c	
24	52	53.1	307	2	O9ZGJ1	ehrlchia c	
25	51	52.0	265	2	O9AF99	ehrlchia c	
26	51	52.0	270	2	O9AF98	ehrlchia c	
27	51	52.0	281	2	O46328	ehrlchia c	
28	51	52.0	281	2	O9S6H1	ehrlchia c	
29	51	52.0	281	2	O9S6H0	ehrlchia c	
30	51	52.0	287	2	O46329	ehrlchia c	
31	51	52.0	287	2	O46331	ehrlchia c	
32	51	52.0	287	2	O9R425	ehrlchia c	
33	51	52.0	290	2	O46324	ehrlchia c	
34	51	52.0	290	2	O46330	ehrlchia c	
35	51	52.0	290	2	O46332	ehrlchia c	
36	51	52.0	290	2	O46333	ehrlchia c	
37	51	52.0	290	2	O9AEU3	ehrlchia c	
38	50	51.0	286	2	O52105	ehrlchia c	
39	50	51.0	330	2	O08347	ehrlchia c	
40	49	50.0	276	2	O9F475	ehrlchia c	
41	48	49.0	439	2	O9KLN9	ehrlchia c	
42	47.5	48.5	612	2	O9F7N4	ehrlchia c	
43	47	48.0	171	2	O9AFAO	ehrlchia c	
44	46	46.9	133	2	O85360	ehrlchia c	
45	46	46.9	283	2	O9F474	ehrlchia c	

ALIGNMENTS

RESULT 1

O52104 PRELIMINARY; PRT; 280 AA.

AC O52104;

DT 01-JUN-1998 (TREMREL. 06, Created)

DT 01-JUN-1998 (TREMREL. 06, Last sequence update)

DT 01-JUN-2001 (TREMREL. 17, Last annotation update)

DE OMP-1C (28 KDA MAJOR SURFACE ANTIGEN-1).

GN OMP-1C

OS Ehrlchia chaffeensis.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Ehrlchieae; Ehrlchia.

OX NCBI_TaxID=945;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ARKANSAS;

RX MEDLINE-98084465; PubMed-9423849;

RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.,

RT "Immunodominant major outer membrane proteins of Ehrlchia chaffeensis are encoded by a polymorphic multigene family.";

RL Infect. Immun. 66:132-139(1998).

RP [2]

RP SEQUENCE OF 238-280 FROM N.A.

RC STRAIN-ARKANSAS;

RX MEDLINE-98321180; PubMed-9647746;

RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,

RA Alleman A.R.;

RT "Molecular characterization of a 28 kDa surface antigen gene family of the tribe Ehrlchieae.";

RL Biochem. Biophys. Res. Commun. 247:636-643(1998).

DR EMBL; U72291; AAC02937.1; "

DR EMBL; AF062761; AAC26717.1; "

DR InterPro: IPR002566; Surface_Ag_msp4.

DR Pfam: PF01617; Surface_Ag_2; 1.

SQ SEQUENCE 280 AA; 30322 MW; BD835E491086DF01 CRC64;

Query Match 100.0%; Score 98; DB 2; Length 280;
Best Local Similarity 100.0%; Pred. No. 4e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNGVSA 18

```
Db 60 NPTVALYGLKQDWNQVSA 77
|||||
RESULT 2
ID O52106 PRELIMINARY; PRT; 278 AA.
AC O52106;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE OMP-LE.
GN OMP-LE.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RL are encoded by a polymorphic multigene family.";
RN Infect. Immun. 66:132-139(1998).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RN Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RN the tribe Ehrlichiae.";
RN Biochem. Biophys. Res. Commun. 247:636-643(1998).
RL EMBL; U72291; AAC02939.1;
DR EMBL; AF062761; AAC26719.1;
DR InterPro: IPR002566; SurfaceAg_msp4.
DR Pfam: PF01617; SurfaceAg_2; 1.
SQ SEQUENCE 278 AA; 30543 MW; E321E3CA259B87FD CRC64;

Query Match 89.8%; Score 88; DB 2; Length 278;
Best Local Similarity 83.3%; Pred. No. 1.7e-06;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNQVSA 18
|||||
Db 60 NPTVALYGLKQDWNQVSA 77

RESULT 3
ID Q9F473 PRELIMINARY; PRT; 280 AA.
AC Q9F473;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE P28-6.
GN P28-6.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=944;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RN antigen.";
RN J. Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
DR EMBL; AF082744; AAG14361.1;
DR InterPro: IPR002566; SurfaceAg_msp4.
DR Pfam: PF01617; SurfaceAg_2; 1.
SQ SEQUENCE 280 AA; 30762 MW; BE284A4B94FE3123 CRC64;

Query Match 63.3%; Score 62; DB 2; Length 280;
Best Local Similarity 66.7%; Pred. No. 0.032;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNQ 15
| | | | |
Db 59 NSTGVGVGLKHDWNG 73

RESULT 4
ID Q9ADV3 PRELIMINARY; PRT; 280 AA.
AC Q9ADV3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN P30-2.
GN P30-2.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RN J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF078553; AAK28699.1;
SQ SEQUENCE 280 AA; 30803 MW; 27238BE1C7E68A91 CRC64;

Query Match 63.3%; Score 62; DB 2; Length 280;
Best Local Similarity 66.7%; Pred. No. 0.032;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNQ 15
| | | | |
Db 59 NSTGVGVGLKHDWNG 73

RESULT 5
ID O85357 PRELIMINARY; PRT; 280 AA.
AC O85357;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
```

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 28 KDA MAJOR SURFACE ANTIGEN-4.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burrige M.J.,
RA Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichiae."
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL: AF062761; AAC36720.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30743 MW; FBB841DAF08EE4DC CRC64;

Query Match 62.2%; Score 61; DB 2; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.046; 2; Indels 0;
Matches 9; Conservative 4; Mismatches 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
Db | | :|||||:
60 NTTIGVFLGKQDWG 74

RESULT 6
Q9ZGJ2 ID Q9ZGJ2 PRELIMINARY; PRT; 288 AA.
AC Q9ZGJ2
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 30-KDA MAJOR OUTER MEMBRANE PROTEIN (P28-8).
GN P30 OR P28-8.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikhisia Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnostics."
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis."
RL Gene 254:245-252(2000).
DR EMBL: AF078553; AAC68667.1; -
DR EMBL: AF082744; AAG14362.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.

DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 288 AA; 31590 MW; 86DCAECB88E9BF5E CRC64;

Query Match 60.2%; Score 59; DB 2; Length 288;
Best Local Similarity 52.9%; Pred. No. 0.1;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVS 17
Db | | :|||||:
60 NTTGTGVLGKQDWGAT 76

RESULT 7
Q9RH35 ID Q9RH35 PRELIMINARY; PRT; 246 AA.
AC Q9RH35
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR (FRAGMENT).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ST. VINCENT;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis."
RL J. Clin. Microbiol. 37:1137-1143(1999).
DR EMBL: AF077735; AAC31548.1; -
DR InterPro: IPR001702; Gram_neg_porin.
DR Pfam: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
FT NON_TER 1
SQ SEQUENCE 246 AA; 26884 MW; C9776392C5129A2F CRC64;

Query Match 59.2%; Score 58; DB 2; Length 246;
Best Local Similarity 60.0%; Pred. No. 0.12;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
Db | | :|||||:
29 NTTAGVFLGKQDWG 43

RESULT 8
O85817 ID O85817 PRELIMINARY; PRT; 276 AA.
AC O85817
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
CC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAPULPA;
RA Yu X.-J., Walker D.H.;
RT "Ehrlichia chaffeensis 28 kDa outer membrane protein."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF077734; AAC31547.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30027 MW; 2FD3698FCF1F60BE CRC64;

Query Match 59.2%; Score 58; DB 2; Length 276;
 Best Local Similarity 60.0%; Pred. No. 0.14;
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
 | | :|:|:|:|:|
 Db 59 NTTAGVGLKQDWG 73

RESULT 9
 O52107 PRELIMINARY; PRT; 280 AA.
 ID O52107
 AC O52107
 DT 01-JUN-1998 (TRENBLrel. 06, Created)
 DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE OMP-IF.
 GN OMP-IF.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
 ON NCBI_TaxID=945;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=ARKANSAS;
 RC MEDLINE=98084465; PubMed=9423849;
 RX Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
 RA "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
 RT are encoded by a polymorphic multigene family.";
 RL Infect. Immun. 66:132-139(1998).
 DR EMBL; U72291; AAC02940.1; -;
 DR InterPro; IPR002566; Surface_Ag_msp4.
 DR Pfam; PF01617; Surface_Ag_2; 1.
 DR CCAAC6C34E2AF393E CRC64;
 SQ SEQUENCE 280 AA; 30731 MW; CCAAC6C34E2AF393E CRC64;

Query Match 59.2%; Score 58; DB 2; Length 280;
 Best Local Similarity 60.0%; Pred. No. 0.14;
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
 | | :|:|:|:|:|
 Db 60 NTTGVGLKQDWG 74

RESULT 10
 O85358 PRELIMINARY; PRT; 276 AA.
 ID O85358
 AC O85358
 DT 01-NOV-1998 (TRENBLrel. 08, Created)
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE 28 KDA MAJOR SURFACE ANTIGEN-5 (FRAGMENT).
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
 ON NCBI_TaxID=945;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=ARKANSAS;
 RC MEDLINE=98321180; PubMed=9647746;
 RX Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burrig M.J.,
 RA Allaman A.R.;
 RA "Molecular Characterization of a 28 kDa surface antigen gene family of
 RT the tribe Ehrlichieae.";
 RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
 DR EMBL; AF062761; AAC26716.1; -;
 DR InterPro; IPR002566; Surface_Ag_msp4.
 DR Pfam; PF01617; Surface_Ag_2; 1.
 DR NON_TER 276 276
 SQ SEQUENCE 276 AA; 29782 MW; A9AB2A92263CA4EA CRC64;

Query Match 58.2%; Score 57; DB 2; Length 276;
 Best Local Similarity 60.0%; Pred. No. 0.21;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
 | | :|:|:|:|:|
 Db 59 NTTGVGLKQDWG 73

RESULT 11
 O9AC19 PRELIMINARY; PRT; 281 AA.
 ID O9AC19
 AC O9AC19
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE MAJOR OUTER MEMBRANE PROTEIN P28.
 GN P28.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
 ON NCBI_TaxID=945;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=ARKANSAS;
 RC MEDLINE=98084465; PubMed=9423849;
 RX Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
 RA "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
 RT are encoded by a polymorphic multigene family.";
 RL Infect. Immun. 66:132-139(1998).
 DR EMBL; U72291; AAC2673.1; -;
 DR SEQUENCE 281 AA; 30343 MW; A99E5F7C4459AA9A CRC64;

Query Match 58.2%; Score 57; DB 2; Length 281;
 Best Local Similarity 60.0%; Pred. No. 0.21;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
 | | :|:|:|:|:|
 Db 59 NTTGVGLKQDWG 73

RESULT 12
 O9AMF6 PRELIMINARY; PRT; 272 AA.
 ID O9AMF6
 AC O9AMF6
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE MAJOR ANTIGENIC PROTEIN MAP1 (FRAGMENT).
 GN MAP1
 OS Cowdria sp. 'South African canine'.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichieae; Cowdria.
 ON NCBI_TaxID=152574;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP Allsopp M.T., Allsopp B.A.;
 RA "A novel Ehrlichia detected in dogs in South Africa.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF325176; AAK14320.1; -;

FT NON_TER 1 1
FT NON_TER 272 272
SQ SEQUENCE 272 AA; 29225 MW; 43CEB95DEDD55BDE CRC64;

Query Match 56.1%; Score 55; DB 2; Length 272;
Best Local Similarity 64.3%; Pred. No. 0.43;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 3 TVALYGLKODWNGV 16
| |::|::|::|::|::|
Db 57 TRAVFGLKKDWDGV 70

RESULT 13
Q46327 PRELIMINARY; PRT; 284 AA.
AC Q46327;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE MAJOR ANTIGENIC PROTEIN.
GN MAP1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Cowdria.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SENEGAL STOCK;
RX MEDLINE=94178956; PubMed=8132352;
RA van Vliet A.H., Jongejan F., van Kleef M., Van der Zeijst B.A.;
RT "Molecular cloning, sequence analysis, and expression of the gene
RT encoding the immunodominant 32-kilodalton protein of Cowdria
RT ruminantium.";
RL Infect. Immun. 62:1451-1456(1994).
DR EMBL; X74250; CAA52309.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 284 AA; 30634 MW; 53228A889D28BEB8 CRC64;

Query Match 56.1%; Score 55; DB 2; Length 284;
Best Local Similarity 64.3%; Pred. No. 0.45;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 3 TVALYGLKODWNGV 16
| |::|::|::|::|::|
Db 63 TRAVFGLKKDWDGV 76

RESULT 14
Q9AFAL PRELIMINARY; PRT; 284 AA.
AC Q9AFAL;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE MAJOR ANTIGENIC PROTEIN 1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Cowdria.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALL-3;
RA Bensaid A., Maillard J.-C., Chantal T., Perez J.-M., Martinez D.;
RT "Cowdria ruminantium major antigenic protein 1 (map1) gene variants
RT are not geographically constrained.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF355200; AAK27216.1; -;
SQ SEQUENCE 284 AA; 30720 MW; B0D3AEB9F9AB09C1 CRC64;

Query Match 56.1%; Score 55; DB 2; Length 284;
Best Local Similarity 64.3%; Pred. No. 0.45;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 3 TVALYGLKODWNGV 16
| |::|::|::|::|::|
Db 63 TRAVFGLKKDWDGV 76

RESULT 15
O85816 PRELIMINARY; PRT; 280 AA.
AC O85816;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-91HE17;
RA Yu X.-J., Walker D.H.;
RT "Ehrlichia chaffeensis 28 kDa outer membrane protein.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077732; AAC31545.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30277 MW; 91C54AC78507A63F CRC64;

Query Match 54.1%; Score 53; DB 2; Length 280;
Best Local Similarity 61.5%; Pred. No. 0.95;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 3 TVALYGLKODWNG 15
| |::|::|::|::|
Db 61 TAGVFLKKDWDG 73

Search completed: March 14, 2002, 09:24:09
Job time: 980 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:24:50 ; Search time 32.82 seconds
(without alignments)
20.109 Million cell updates/sec

Title: US-09-765-739A-4
Perfect score: 98
Sequence: 1 NPTVALYGLKQDWNGVSA 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %	Score	Match	Length	DB	ID	Description
1	47	48.0	341	1	Y33B_MYCPN	P75302	mycoplasma
2	43	43.9	918	1	QA1S_NEUCR	P11637	neurospora
3	42.5	43.4	1803	1	YJL3_YEAST	P47024	saccharomyc
4	42	42.9	311	1	CDX2_HUMAN	Q99626	homo sapien
5	42	42.9	1023	1	TSCC_PSEAM	P55019	pseudopleur
6	41	41.8	184	1	VP50_BPAPS	Q911p8	bacterioph
7	41	41.8	290	1	AKOE_SYNY3	P74591	synecocyst
8	41	41.8	459	1	Y030_NPVOP	O10293	orgyia pseu
9	41	41.8	544	1	PYRG_AZOBR	P28595	azospirillu
10	41	41.8	874	1	SVV_STRCO	O06851	streptomyce
11	40.5	41.3	354	1	CARA_METJA	Q58425	methanococ
12	40	40.8	311	1	CDX2_MOUSE	P43241	mus musculu
13	40	40.8	615	1	UPAP_EMENI	Q07307	emericeila
14	40	40.8	734	1	PURL_ZYMMO	Q9req6	zymomonas m
15	40	40.8	843	1	PULA_THEMA	O33840	thermotoga
16	40	40.8	991	1	DHPI_SCHPO	P40848	schizosacch
17	40	40.8	1060	1	NKCL_MANSE	Q25479	manduca sex
18	39.5	40.3	726	1	ADB2_YEAST	P36000	saccharomyc
19	39	39.8	228	1	YEBB_METEX	Q49116	methylobact
20	39	39.8	237	1	DCOP_LACLC	P50924	lactococcus
21	39	39.8	263	1	OMPK_VIBPA	P51002	vibrio para
22	39	39.8	286	1	RIP1_MOMCH	P16094	momordica c
23	39	39.8	313	1	CDX2_MESAU	Q04649	mesocricetu
24	39	39.8	329	1	Y830_METTH	O26918	methanobact
25	39	39.8	353	1	SWB_BORBU	O51038	borrelia bu
26	39	39.8	493	1	VG10_BPMD2	O64203	mycobacteri
27	39	39.8	589	1	VP40_SCMVC	P16046	simian cyto
28	39	39.8	643	1	VP40_HSV2	P52369	equine herp
29	39	39.8	671	1	ALYS_ENTFA	P37710	enterococcu
30	39	39.8	732	1	EF2_PYRAB	Q9V128	pyrococcus
31	39	39.8	732	1	EF2_PYRHO	O59521	pyrococcus
32	39	39.8	868	1	PRTT_PORGI	P43158	porphyromon
33	39	39.8	1191	1	NKCL_SQUAC	P55013	squalus aca

34	39	39.8	2907	1	FBN2_MOUSE	Q61555	mus musculu
35	39	39.8	2911	1	FBN2_HUMAN	P35556	homo sapien
36	38.5	39.3	192	1	Y544_BUCAI	P57609	buchnera ap
37	38.5	39.3	347	1	YCEG_HAEIN	P44720	haemophilus
38	38.5	39.3	393	1	VGLD_HSV2	P03172	herpes simp
39	38.5	39.3	396	1	ODP2_BUCAI	P57302	buchnera ap
40	38.5	39.3	643	1	PGT_HUMAN	Q92959	homo sapien
41	38	38.8	55	1	ANTA_HIRME	P80302	hirudo medi
42	38	38.8	137	1	VIT_STRPU	P19615	strongyloce
43	38	38.8	280	1	ETA_STAAR	P09331	staphylococ
44	38	38.8	300	1	DAPA_RHIME	O69782	rhizobium m
45	38	38.8	355	1	GLN1_PEA	P08282	pisum sativ

ALIGNMENTS

RESULT 1
Y33B_MYCPN
ID Y33B_MYCPN STANDARD; PRT; 341 AA.
AC P75302;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE HYPOTHETICAL PROTEIN MG335.2 HOMOLOG (P01_ORF341).
GN MPN483 OR MP359.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R.; Hilbert H.; Plagens H.; Pirkl E.; Li B.-C.;
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC STRONG, TO M.GENITALIUM MG335.2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL: AF000034; AAB96007.1; -
CC InterPro: IPR001173; Glycosyltransf_2.
CC Pfam: PF00535; Glycosyltransf_2; 1.
CC KW Hypothetical protein; Transferase; Glycosyltransferase;
CC Complete proteome.
SQ SEQUENCE 341 AA; 40414 MW; C209F50D714CB3D0 CRC64;

Query Match 48.0%; Score 47; DB 1; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LKQDWNGV 16
Db 26 LKQDWNGV 33

RESULT 2
QA1S_NEUCR
ID QA1S_NEUCR STANDARD; PRT; 918 AA.
AC P11637;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)

```

DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE QUINATE REPRESSOR.
GN QA-15.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=89233848; PubMed=2525625;
RA Geever R.F., Huiet L., Baum J.A., Patel V.B.,
RA Rutledge B.J., Case W.E., Giles N.H.;
RT "DNA sequence, organization and regulation of the qa gene cluster of
RT Neurospora crassa.";
RL J. Mol. Biol. 207:15-34(1989).
CC -1- FUNCTION: REPRESSOR FOR ENZYMES AND PROTEINS OF QUINATE
CC METABOLISM.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X14603; CAA32753.1;
CC PIR: S04255; S04255.
CC PIR: E31277; E31277.
CC InterPro: IPR001381; DHQuinase_I.
CC InterPro: IPR000623; Shik_kinase.
CC InterPro: IPR002907; Shikimate_DH.
CC Pfam: PF01487; DHQuinase_I; 1.
CC Pfam: PF01488; Shikimate_DH; 1.
CC Pfam: PF01202; SKI; 1.
CC Quinate metabolism; Transcription regulation; Repressor; DNA-binding.
KW Quinate metabolism; Transcription regulation; Repressor; DNA-binding.
SQ SEQUENCE 918 AA; 100580 MW; 67EDA399CBF098B2 CRC64;

Query Match 43.9%; Score 43; DB 1; Length 918;
Best Local Similarity 57.1%; Pred. No. 23;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 5 ALYGLKODWNGVSA 18
   |||| | | | |
DB 684 ALYXGTDWIGIRA 697

RESULT 3
YJL3_YEAST
ID YJL3_YEAST STANDARD; PRT: 1803 AA.
AC P47024; P87192;
DT 01-FEB-1996 (Rel. 33, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSPOSON TY4 207.7 KDA HYPOTHETICAL PROTEIN.
GN TY4B OR YJL13W OR J0780.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97103775; PubMed=8948101;
RA Cziepluch C., Kordes E., Pujol A., Jauniaux J.-C.;
RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
RT reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
RT SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, CCF7, GZF3, two tRNA genes,
RT three remnant delta elements and a Ty4 transposon.";
RL Yeast 12:1471-1474(1996).
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X14603; CAA32753.1;
CC PIR: S04255; S04255.
CC PIR: E31277; E31277.
CC InterPro: IPR001381; DHQuinase_I.
CC InterPro: IPR000623; Shik_kinase.
CC InterPro: IPR002907; Shikimate_DH.
CC Pfam: PF01487; DHQuinase_I; 1.
CC Pfam: PF01488; Shikimate_DH; 1.
CC Pfam: PF01202; SKI; 1.
CC Quinate metabolism; Transcription regulation; Repressor; DNA-binding.
KW Quinate metabolism; Transcription regulation; Repressor; DNA-binding.
SQ SEQUENCE 918 AA; 100580 MW; 67EDA399CBF098B2 CRC64;

Query Match 43.4%; Score 42.5; DB 1; Length 1803;
Best Local Similarity 69.2%; Pred. No. 57;
Matches 9; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

QY 5 ALYGLKQSPKEWN 1427
   ||||| | | |
DB 1415 ALYGLKQSPKEWN 1427

RESULT 4
CDX2_HUMAN
ID CDX2_HUMAN STANDARD; PRT: 311 AA.
AC Q9626; Q00503;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HOMEBOX PROTEIN CDX-2 (CAUDAL-TYPE HOMEBOX PROTEIN 2) (CDX-3).
GN CDX2 OR CDX3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon carcinoma;
RX MEDLINE=97188282; PubMed=9036867;
RA Mallo G.V., Rechreche H., Frigerio J.M., Rocha D., Zweibaum A.,
RA Lacasa M., Jordan B.R., Dusetti N.J., Dagorn J.C., Iovanna J.L.;
RT "Molecular cloning, sequencing and expression of the mRNA encoding
RT human Cdx1 and Cdx2 homeobox. Down-regulation of Cdx1 and Cdx2 mRNA
RT expression during colorectal carcinogenesis.";
RL Int. J. Cancer 74:35-44(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98120622; PubMed=9459001;
RA Drummond F.J., Putt W., Fox M., Edwards Y.H.;
RT "Cloning and chromosome assignment of the human CDX2 gene.";
RL Ann. Hum. Genet. 61:393-400(1997).
CC -1- FUNCTION: MAY BE NECESSARY FOR SOME GENERAL ASPECT OF COLONIC
CC EPITHELIAL PHENOTYPE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR
CC -1- SIMILARITY: BELONGS TO THE CAUDAL FAMILY OF HOMEBOX PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U51096; AAB40603.1;
CC EMBL: Y13709; CAA74038.1;
CC HSSP: P02833; 1SAN.
CC MIN: 600297;

```


DR InterPro: IPR000047; HTH_repressr.
 DR InterPro: IPR001356; Homeobox.
 DR Pfam: PF00046; homeobox; 1.
 DR PRINTS: PR00024; HOMEBOX.
 DR PRINTS: PR00031; HTHREPRESSR.
 DR SMART: SM00389; HOX; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS00711; HOMEBOX_2; 1.
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
 FT DOMAIN 46 51
 FT DOMAIN 84 91
 FT DNA_BIND 184 243
 FT DOMAIN 248 254
 FT DOMAIN 255 268
 FT CONFLICT 52 52 Q -> AA (IN REF. 2).
 FT CONFLICT 87 87 A -> AA (IN REF. 2).
 FT CONFLICT 93 93 A -> G (IN REF. 2).
 SQ SEQUENCE 311 AA; 33438 MW; C2FEDEF1089D2367 CRC64;

 Query Match 42.9%; Score 42; DB 1; Length 311;
 Best Local Similarity 61.5%; Pred. No. 11;
 Matches 8; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

 QY 5 ALYG--LKQDNG 15
 | | | | |
 Db 66 AAYGAPLREDNG 78

 RESULT 5
 TSSC_PSEAM
 ID TSSC_PSEAM STANDARD; PRT; 1023 AA.
 AC P55019;
 DT 01-OCT-1996 (Rel. 34; Created)
 DT 01-OCT-1996 (Rel. 34; Last sequence update)
 DT 01-NOV-1997 (Rel. 35; Last annotation update)
 DE THIAZIDE-SENSITIVE SODIUM-CHLORIDE COTRANSPORTER (NA-CL SYMPORTER).
 GN SLC12A3 OR TSC.
 OS Pseudopleuronectes americanus (winter flounder).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
 CC Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.
 OX NCBI_TaxID=8265;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bladder urothelium;
 RX MEDLINE=93219361; PubMed=8464884;
 RA Gamba G., Saltzberg S.N., Lombardi M., Miyanoishi A., Lytton J.,
 RA Hediger M.A., Brenner B.M., Hebert S.C.;
 RT "Primary structure and functional expression of a cDNA encoding the
 RT thiazide-sensitive, electroneutral sodium-chloride cotransporter.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:2749-2753(1993).
 CC -!- FUNCTION: ELECTRICALLY SILENT TRANSPORTER SYSTEM WHICH IS A
 CC MEDIATOR OF SODIUM AND CHLORIDE REABSORPTION.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: URINARY BLADDER.
 CC -!- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L11615; AAA49272.1;
 DR InterPro: IPR002293; AA_rel_permease.1.
 DR InterPro: IPR002027; Amino_acid_permease.
 DR InterPro: IPR002948; NaCl_trnsporr.
 DR Pfam; PF00324; aa_permeases; 1.
 DR PRINTS: PR01230; NACLTRNSPORT.

KW Transport; Transmembrane; Glycoprotein.
 FT DOMAIN 1 132
 FT TRANSMEM 133 153
 FT DOMAIN 154 163
 FT TRANSMEM 164 184
 FT DOMAIN 185 215
 FT TRANSMEM 216 236
 FT DOMAIN 237 258
 FT TRANSMEM 259 279
 FT DOMAIN 280 283
 FT TRANSMEM 284 304
 FT DOMAIN 305 336
 FT TRANSMEM 337 357
 FT DOMAIN 358 374
 FT TRANSMEM 375 395
 FT DOMAIN 396 451
 FT TRANSMEM 452 472
 FT DOMAIN 473 510
 FT TRANSMEM 511 531
 FT DOMAIN 532 576
 FT TRANSMEM 577 597
 FT DOMAIN 598 745
 FT TRANSMEM 746 766
 FT DOMAIN 767 830
 FT TRANSMEM 831 851
 FT DOMAIN 852 1023
 FT TRANSMEM 77 82
 FT DOMAIN 117 122
 FT TRANSMEM 407 412
 FT DOMAIN 933 937
 FT CARBOHYD 403 403
 FT CARBOHYD 414 414
 FT CARBOHYD 432 432
 FT CARBOHYD 543 543
 SQ SEQUENCE 1023 AA; 112340 MW; 69AE2D53B8F84D89 CRC64;

 Query Match 42.9%; Score 42; DB 1; Length 1023;
 Best Local Similarity 58.3%; Pred. No. 39;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

 QY 2 PIVALYGLKQDW 13
 | | | | |
 Db 727 PNVLLMGFKKW 738

 RESULT 6
 VP50_BPAPS
 ID VP50_BPAPS STANDARD; PRT; 184 AA.
 AC Q9TIP8;
 DT 20-AUG-2001 (Rel. 40; Created)
 DT 20-AUG-2001 (Rel. 40; Last sequence update)
 DT 20-AUG-2001 (Rel. 40; Last annotation update)
 DE PUTATIVE PROTEIN P50.
 GN 50.
 OS Bacteriophage APSE-1.
 OC Viruses.
 OX NCBI_TaxID=106199;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99420383; PubMed=10489345;
 RA van der Wilk F., Dullemans A.M., Verbeek M., van den Heuvel J.F.J.M.;
 RT "Isolation and characterization of APSE-1, a bacteriophage infecting
 RT the secondary endosymbiont of acyrthosiphon pisum.";
 RL Virology 262:104-113(1999).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC EMBL; AF157835; AAF03993.1;
DR Hypothetical protein.
SQ SEQUENCE 184 AA; 20416 MW; EB705FA272B6768F CRC64;

Query Match
Best Local Similarity 41.8%; Score 41; DB 1; Length 184;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWMNGVSA 18
DB 128 NATVEFYGDNDNGKGVSA 145

RESULT 7
ID AROELSYNY3 STANDARD; PRT; 290 AA.
AC 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE SHIKIMATE 5 (Rel. 40, Last annotation update)
GN AROE OR SLR1559.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803 II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- CATALYTIC ACTIVITY: SHIKIMATE + NADP(+) = 5-DEHYDROSHIKIMATE +
CC NADPH.
CC -!- PATHWAY: FOURTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN THE
CC BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).
CC -!- SIMILARITY: BELONGS TO THE SHIKIMATE DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D90916; BAA18699.1;
DR InterPro; IPR002907; Shikimate_DH.
DR Pfam; PF01488; Shikimate_DH; 1.
KW Aromatic amino acid biosynthesis; Oxidoreductase; NADP;
KW Complete proteome.
SQ SEQUENCE 290 AA; 31099 MW; 8A2D38EE5D57B303 CRC64;

Query Match
Best Local Similarity 41.8%; Score 41; DB 1; Length 290;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 6 LYGLKQDWMNGVSA 18
DB 116 LLELKQDWMNGVSA 128

RESULT 8
ID Y030_NPVOP STANDARD; PRT; 459 AA.

```

```

AC 010293;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 52.7 KDA PROTEIN (ORF38).
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolydnavirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohrmann G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome.";
RL Virology 229:381-399(1997).
CC -!- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U75930; AAC59037.1;
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 459 AA; 52731 MW; DAAFBDB8C6628F596 CRC64;

Query Match
Best Local Similarity 41.8%; Score 41; DB 1; Length 459;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 VALYGLKQDWMNGVSA 18
DB 56 LSAFGDKIDWNEVSA 70

RESULT 9
ID PYRG_AZOBR STANDARD; PRT; 544 AA.
AC P28595;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE CTP SYNTHASE (EC 6.3.4.2) (UTP--AMMONIA LIGASE) (CTP SYNTHETASE).
GN PYRG.
OS Azospirillum brasilense.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
OC Azospirillum.
OX NCBI_TaxID=192;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 29145 / SP7;
RA MEDLINE=94186025; PubMed=8138139;
RA Zimmer W., Hundschaen B.;
RT "Identification and sequencing of pyrG, the CTP synthetase gene of
RT Azospirillum brasilense Sp7.";
RL FEMS Microbiol. Lett. 115:273-278(1994).
CC -!- CATALYTIC ACTIVITY: ATP + UTP + GLUTAMINE = ADP + ORTHOPHOSPHATE +
CC CTP (AMMONIA CAN REPLACE GLUTAMINE).
CC -!- ENZYME REGULATION: ACTIVATED BY GTP AND INHIBITED BY CTP.
CC -!- PATHWAY: LAST STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -!- SIMILARITY: TO OTHER SPECIES CTP SYNTHASE.
CC -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-1 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

EMBL; X67216; CAA47656.1; -
PIR; S25101; S25101.
InterPro; IPR000991; GATase_1.
Pfam; PF00117; GATase; 1.
PROSITE; PS00442; GATASE_TYPE_I; 1.
Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase.
ACT_SITE 380 380 GATASE (BY SIMILARITY).
ACT_SITE 516 516 GATASE (BY SIMILARITY).
ACT_SITE 518 518 GATASE (BY SIMILARITY).
SEQUENCE 544 AA; 59908 MW; 13AA1A8B8EB6C5DB CRC64;

Query Match 41.8%; Score 41; DB 1; Length 544;
Best Local Similarity 53.8%; Pred. NO. 30;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 NPTVALYGLKQDW 13
DB 410 NPVGLGLMTW 422

RESULT 10
ID SVV_STRCO STANDARD; PRT; 874 AA.

AC O06851; Q9L1G3;
DT 15-JUL-1998 (Rel. 36, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VALYL-TRNA SYNTHETASE (EC 6.1.1.9) (VALINE--TRNA LIGASE) (VALRS).
GN VALS OR SCC88.26C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D., Harris D., Cerdano A.M., Parkhill J., Barrell B.G.,
RA Rajadream M.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 616-873 FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=98164371; PubMed=9503623;

RA Burger A., Brandt B., Suesstrunk U., Thompson C.J., Wohlleben W.;
RT "Analysis of a Streptomyces coelicolor A3(2) locus containing the
RT nucleoside diphosphate kinase (ndk) and folylpolyglutamate synthetase
RT (folC) genes.";
RL FEMS Microbiol. Lett. 159:283-291(1998).
CC -!- CATALYTIC ACTIVITY: ATP + L-VALINE + TRNA(VAL) = AMP +
CC PYROPHOSPHATE + L-VALYL-TRNA(VAL).
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

EMBL; AL139298; CAB75396.1; -
EMBL; Y13070; CAA73510.1; -
InterPro; IPR002300; tRNA_synth_1a.
InterPro; IPR001412; tRNA-synt_1.
InterPro; IPR002303; tRNA-synt_val.

Pfam; PF00133; tRNA-synt_1; 1.
PRINTS; PR00986; TRNASYNTHVAL.
PROSITE; PS00178; AA_TRNA_LIGASE_I; PARTIAL.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT SITE 57 67 "HIGH" REGION.
FT SITE 531 535 "KMSKS" REGION.
FT BINDING 534 534 ATP (BY SIMILARITY).
FT CONFLICT 638 639 KL -> NV (IN REF. 1).
SQ SEQUENCE 874 AA; 97566 MW; AD11E4871786BAA0 CRC64;

Query Match 41.8%; Score 41; DB 1; Length 874;
Best Local Similarity 53.8%; Pred. NO. 48;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 TVALYGLKQDWNG 15
DB 517 TIALHGMRDQNG 529

RESULT 11
ID CARA_METJA STANDARD; PRT; 354 AA.

AC Q58425;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CARBAMOYL-PHOSPHATE SYNTHASE SMALL CHAIN (EC 6.3.5.5) (CARBAMOYL-
DE PHOSPHATE SYNTHETASE GLUTAMINE CHAIN).
GN CARA OR MJ1019.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.S.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: 2 ATP + GLUTAMINE + CO(2) + H(2)O = 2 ADP +
CC ORTHOPHOSPHATE + GLUTAMATE + CARBAMOYL PHOSPHATE.
CC -!- PATHWAY: INVOLVED IN BOTH ARGinine AND PYrimidine BIOSYNTHESIS.
CC -!- SUBUNIT: COMPOSED OF TWO CHAINS; THE SMALL (OR GLUTAMINE) CHAIN
CC PROMOTES THE HYDROLYSIS OF GLUTAMINE TO AMMONIA, WHICH IS USED
CC BY THE LARGE (OR AMMONIA) CHAIN TO SYNTHESIZE CARBAMOYL PHOSPHATE.
CC -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-1 GLUTAMINE
CC AMIDOTRANSFERASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

EMBL; U67544; AAB99021.1; -
HSPSP; P00907; 1JDB.
TIGR; MJ1019;
InterPro; IPR001317; CPS_GATase.
InterPro; IPR002474; CPSase_sm_chain.
InterPro; IPR000991; GATase_1.
Pfam; PF00988; CPSase_sm_chain; 1.

```

DR Pfam: PF00117; GATase; 1.
DR PRINTS: PR00096; GATASE.
DR PRINTS: PR00099; CPSGATASE.
DR PROSITE: PS00442; GATASE_TYPE_1; FALSE_NEG.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase;
KW Complete proteome.
FT DOMAIN 1 ? CPSASE.
FT DOMAIN 2 ? 354 GLUTAMINE AMIDOTRANSFERASE.
FT ACT_SITE 250 250 GATASE (BY SIMILARITY).
SQ SEQUENCE 354 AA; 39920 MW; 4DBF07776CA65F9F CRC64;

Query Match 41.3%; Score 40.5; DB 1; Length 354;
Best Local Similarity 40.8%; Pred. No. 23;
Matches 8; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

QY 2 PTVALYGLQDW---NGVSA 18
DB 56 PLEGNYGVKRWFSDDGKRA 75
|||||:||||:|:|

RESULT 12
CDX2_MOUSE STANDARD; PRT; 311 AA.
ID CDX2_MOUSE AC P43241;
DT 01-NOV-1995 (Rel. 32; Created)
DT 01-NOV-1995 (Rel. 32; Last sequence update)
DT 15-JUL-1998 (Rel. 36; Last annotation update)
DE HOMEBOX PROTEIN CDX-2 (CAUDAL-TYPE HOMEBOX PROTEIN 2).
GN CDX2 OR CDX-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C.
RX MEDLINE=94253086; PubMed=7910823;
RA James R.J., Exler T., Kazenwadel J.;
RT "Structure of the murine homeobox gene cdx-2. Expression in embryonic
RT and adult intestinal epithelium."
RL J. Biol. Chem. 269:15229-15237(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Small intestine;
RX MEDLINE=95021263; PubMed=7935448;
RA Suh E., Chen L., Taylor J., Traber P.G.;
RT "A homeodomain protein related to caudal regulates intestine-specific
RT gene transcription."
RL Mol. Cell. Biol. 14:7340-7351(1994).
RN [3]
RP SEQUENCE OF 204-229 FROM N.A.
RX MEDLINE=91131633; PubMed=1671571;
RA James R.J., Kazenwadel J.;
RT "Homeobox gene expression in the intestinal epithelium of adult
RT mice."
RL J. Biol. Chem. 266:3246-3251(1991).
CC -!- FUNCTION: MAY BE NECESSARY FOR SOME GENERAL ASPECT OF COLONIC
CC -!- EPITHELIAL PHENOTYPE.
CC -!- TISSUE SPECIFICITY: INTESTINE; EXPRESSED SPECIFICALLY IN GUT
CC EPITHELIUM WHERE IT IS NOT RESTRICTED TO A PARTICULAR CELL
CC LINEAGE. ABUNDANT EXPRESSION IS SEEN IN THE PROXIMAL COLON WITH
CC SLIGHTLY LOWER LEVELS IN DISTAL COLON. EXPRESSION IN THE PROXIMAL
CC COLON IS NOT RESTRICTED EITHER TO A PARTICULAR CELL LINEAGE OR
CC STAGE OF DIFFERENTIATION WHILE IN THE DISTAL COLON IT IS MORE
CC ABUNDANT IN THE DIFFERENTIATED CELLS TOWARDS THE TOP OF THE CRYPT.
CC -!- SIMILARITY: BELONGS TO THE CAUDAL FAMILY OF HOMEBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X71807; CAA50681.1; -

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00454; AAA19645.1; -
CC EMBL; S74520; AAB32251.1; -
CC HSSP: P02833; ISAN.
CC TRANSFAC: T02002; -
CC MGD: MGI:88361; Cdx2.
CC InterPro: IPR000047; HTH_Repressr.
CC InterPro: IPR001356; Homeobox.
CC Pfam: PF00046; homeobox; 1.
CC PRINTS: PS00024; HOMEBOX.
CC PRINTS: PS00031; HTHREPRESSR.
CC SMART: SM00389; HOX; 1..XX; 1.
CC PROSITE: PS00021; HOMEBOX_1; 1.
CC PROSITE: PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DOMAIN 46 53 POLY-ALA.
FT DOMAIN 85 92 POLY-ALA.
FT DNA_BIND 185 244 HOMEBOX.
FT DOMAIN 247 257 POLY-GLN.
FT CONFLICT 69 69 Y -> H (IN REF. 2).
SQ SEQUENCE 311 AA; 33476 MW; 71FFC4C263462FF3 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 311;
Best Local Similarity 56.2%; Pred. No. 24;
Matches 9; Conservative 2; Mismatches 1; Indels 4; Gaps 2;

QY 2 PTVALYGLQDWNG 15
DB 66 PTA--YGAPLREDWNG 79
|||||:|:|:|:|

RESULT 13
UAPA_EMENI STANDARD; PRT; 615 AA.
ID UAPA_EMENI AC Q07307;
DT 01-OCT-1994 (Rel. 30; Created)
DT 15-JUL-1999 (Rel. 38; Last sequence update)
DT 15-JUL-1999 (Rel. 38; Last annotation update)
DE URIC ACID-XANTHINE PERMEASE (UAPA TRANSPORTER).
GN UAPA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94043131; PubMed=8226862;
RA Gorfinkel L., Diallinas G., Scazzocchio C.;
RT "Sequence and regulation of the uapA gene encoding a uric acid-
RT xanthine permease in the fungus Aspergillus nidulans."
RL J. Biol. Chem. 268:23376-23381(1993).
RN [2]
RP REVISTONS.
RA Diallinas G.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- INDUCTION: INDUCIBLE BY 2-THIOURIC ACID, AND HIGHLY REPRESSIBLE BY
CC AMMONIUM.
CC -!- SIMILARITY: BELONGS TO THE XANTHINE/URACIL PERMEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X71807; CAA50681.1; -

```

DR PIR: S36031; S36031.
 DR InterPro; IPR000444; xan_ur_permease.
 DR Pfam; PF00860; xan_ur_permease; 1.
 DR PROSITE; PS01116; XANTH_URACIL_PERMEASE; 1.
 KW Transport; Transmembrane; Glycoprotein.
 FT TRANSMEM 118 138 POTENTIAL.
 FT TRANSMEM 152 172 POTENTIAL.
 FT TRANSMEM 182 202 POTENTIAL.
 FT TRANSMEM 229 250 POTENTIAL.
 FT TRANSMEM 258 278 POTENTIAL.
 FT TRANSMEM 305 325 POTENTIAL.
 FT TRANSMEM 337 356 POTENTIAL.
 FT TRANSMEM 379 402 POTENTIAL.
 FT TRANSMEM 468 488 POTENTIAL.
 FT TRANSMEM 492 512 POTENTIAL.
 FT TRANSMEM 523 543 POTENTIAL.
 FT TRANSMEM 563 583 POTENTIAL.
 FT CARBOHYD 10 10 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 615 AA; 65453 MW; E56A984D956897E7 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 615;
 Best Local Similarity 43.5%; Pred. No. 49;
 Matches 10; Conservative 2; Mismatches 5; Indels 6; Gaps 1;

QY 2 PTVALYGLK-----QDMNGVSA 18
 ||||| : :|||
 DB 266 PTVMLIGLSLTGTFKDWAGGSA 288

RESULT 14
 PURL_ZYMMO STANDARD; PRT; 734 AA.
 ID PURL_ZYMMO
 RC STRAIN-ATCC 31821 / ZM4 / CP4;
 RA Um H.W., Kang H.S.;
 CC Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: ATP + 5'-PHOSPHORIBOSYLFORMYLGLYCINAMIDE +
 CC L-GLUTAMINE + H(2)O = ADP + ORTHOPHOSPHATE + 5'-PHOSPHORIBOSYL-
 CC FORMYLGLYCINAMINE + L-GLUTAMATE.
 CC -!- PATHWAY: DE NOVO PURINE BIOSYNTHESIS; FOURTH STEP.
 CC -!- SUBUNIT: HETERODIMER OF TWO SUBUNITS, PURQ AND PURL.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE FGAMS FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF213822; AAF23789.1; -.
 DR InterPro; IPR000728; AIRS-related.
 DR Pfam; PF00586; AIRS; 2.
 KW Purine biosynthesis; Ligase; ATP-binding.
 FT NP_BIND 106 117
 SQ SEQUENCE 734 AA; 77679 MW; B771635E0F66A166 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 734;
 Best Local Similarity 58.3%; Pred. No. 59;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PTVALYGLKQDW 13
 ||: |||||
 DB 550 PTIGGVGLQDW 561

RESULT 15
 PULA_THEMA STANDARD; PRT; 843 AA.
 ID PULA_THEMA
 AC O33840;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PULLULANASE PRECURSOR (EC 3.2.1.41) (ALPHA-DEXTRIN ENDO-1,6-ALPHA-
 DE GLUCOSIDASE) (PULLULAN 6-GLUCANOHYDROLASE).
 GN PULA OR TM1845.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID-2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MSB8 / DSM 3109;
 RX MEDLINE-98115241; PubMed-9453151;
 RA Bibel M., Brettl C., Gossler U., Kriegshaeuser G., Liebl W.;
 RT "Isolation and analysis of genes for amylolytic enzymes of the
 RT hyperthermophilic bacterium Thermotoga maritima.";
 RL FEMS Microbiol. Lett. 158:9-15(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MSB8 / DSM 3109;
 RX MEDLINE-99287316; PubMed-10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima.";
 RL Nature 398:323-329(1999).
 CC -!- CATALYTIC ACTIVITY: STARCH-DEBRANCHING ENZYME, HYDROLYSES
 CC (1-6)-ALPHA-GLUCOSIDIC LINKAGES IN PULLULAN AND STARCH TO
 CC FORM MALTOSE.
 CC -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AJ001087; CAA04522.1; -.
 DR EMBL; AE001821; AAD36907.1; -.
 DR TIGR; TM1845; -.
 DR InterPro; IPR000461; Alpha_amylase.
 DR Pfam; PF00128; alpha-amylase; 1.
 KW Hydrolase; Glycosidase; Signal; Complete proteome.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 843 PULLULANASE.
 FT ACT_SITE 535 535 BY SIMILARITY.
 FT ACT_SITE 564 564 BY SIMILARITY.
 FT ACT_SITE 652 652 BY SIMILARITY.
 SQ SEQUENCE 843 AA; 96261 MW; C42DDE233D54FE77 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 843;
 Best Local Similarity 44.4%; Pred. No. 68;

us-09-765-739a-4.rsp

Matches	8;	Conservative	3;	Mismatches	5;	Indels	2;	Gaps	1;
---------	----	--------------	----	------------	----	--------	----	------	----

QY 1 NPTVALYGLKQDWNGVSA 18
:||: ||| : | | |
Db 556 DPTIILYG--EPWGGWGA 571

Search completed: March 14, 2002, 09:24:51
Job time: 907 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:10:42 ; Search time 56.96 Seconds
(without alignments)
26.747 Million cell updates/sec

Title: US-09-765-739A-3

Perfect score: 107

Sequence: 1 NTTVGVGFLKQNDGSAISN 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*

- 1: pir1:*
- 2: pir2:*
- 3: pir3:*
- 4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	276	JE0218	28k surface antige
2	91	85.0	280	JE0217	28k surface antige
3	73	68.2	286	JE0219	28k surface antige
4	61.5	57.5	133	JE0221	28k surface antige
5	60	56.1	278	JE0216	28k surface antige
6	53	49.5	284	I40882	major antigenic pr
7	46	43.0	299	T23932	hypothetical prote
8	46	43.0	1004	JH0470	Na+/K+-exchanging
9	46	43.0	1034	J76134	hypothetical prote
10	45	42.1	160	A75466	2-demethylmenaquin
11	45	42.1	300	S60558	envelope polyprote
12	45	42.1	300	S60526	envelope polyprote
13	45	42.1	490	G85354	hypothetical prote
14	45	42.1	540	S54586	probable membrane
15	44	41.1	1037	A56594	Na+/K+-exchanging
16	44	41.1	1649	C86822	hypothetical prote
17	43.5	40.7	165	G83166	hypothetical prote
18	43	40.2	86	S28126	gas-vesicle protei
19	43	40.2	384	T40502	hypothetical prote
20	43	40.2	422	S73667	adhesin P1 precurs
21	43	40.2	428	S73379	adhesin P1 precurs
22	43	40.2	438	S73952	adhesin P1 precurs
23	43	40.2	547	E81017	ABC transporter, A
24	43	40.2	648	1 P3BPF6	P3 protein - phage
25	42.5	39.7	534	C82096	aminoacyl-histidin
26	42.5	39.7	1538	S73296	glutamate synthase
27	42	39.3	277	S54501	probable membrane
28	42	39.3	297	T21632	hypothetical prote
29	42	39.3	467	S30839	UTR2 protein - yea

30	42	39.3	631	2	T20037	hypothetical prote
31	42	39.3	1038	1	S03632	Na+/K+-exchanging
32	41.5	38.8	540	2	C84744	probable PPIg-like
33	41	38.3	173	2	S65999	streptothricine ac
34	41	38.3	233	2	T06915	ribosomal protein
35	41	38.3	264	2	T13743	hypothetical prote
36	41	38.3	282	2	F64695	hypothetical prote
37	41	38.3	282	2	H71823	biotin synthetase
38	41	38.3	330	2	S76408	hypothetical prote
39	41	38.3	331	2	C72239	hypothetical prote
40	41	38.3	357	2	T25499	hypothetical prote
41	41	38.3	363	1	SYM502	(2'-5')oligo(A) sy
42	41	38.3	364	1	SYHU16	(2'-5')oligo(A) sy
43	41	38.3	377	2	F69008	acetyltransferase
44	41	38.3	400	1	SYHU18	(2'-5')oligo(A) sy
45	41	38.3	414	1	SYM503	(2'-5')oligo(A) sy

ALIGNMENTS

RESULT 1
JE0218
28k surface antigen 5 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0218
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0218
A:Molecule type: DNA
A:Residues: 1-276 <RED>
A:Cross-references: GB:AF062761

Query Match 100.0%; Score 107; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 1.2e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTVGVGFLKQNDGSAISN 20
|||||
DB 59 NTTVGVGFLKQNDGSAISN 78

RESULT 2
JE0217
28k surface antigen 4 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0217
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0217
A:Molecule type: DNA
A:Residues: 1-280 <RED>
A:Cross-references: GB:AF062761

Query Match 85.0%; Score 91; DB 2; Length 280;
Best Local Similarity 84.2%; Pred. No. 4.6e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTVGVGFLKQNDGSAIS 19
|||||
DB 60 NTTIGVGFLKQNDGSAIS 78

A:Accession: JH0470
A:Molecule type: mRNA
A:Residues: 1-1004 <MAC>
A:Cross-references: EMBL:X56650; NID:g10933; PIDN:CAA39972.1; PID:g10934
C:Superfamily: Na⁺/K⁺-transporting ATPase alpha chain; ATPase nucleotide-binding domain
C:Keywords: ATP; heterodimer; hydrolase; ion transport; phosphoprotein; potassium transp
F:2-1004/Product: Na⁺/K⁺-transporting ATPase alpha chain #status predicted <MAT>
F:2-75/Domain: intracellular #status predicted <INT1>
F:76-97/Domain: transmembrane #status predicted <TM1>
F:111-130/Domain: transmembrane #status predicted <TM2>
F:131-271/Domain: intracellular #status predicted <INT2>
F:272-296/Domain: transmembrane #status predicted <TM3>
F:301-329/Domain: transmembrane #status predicted <TM4>
F:330-767/Domain: intracellular #status predicted <INT>
F:568-764/Domain: ATPase nucleotide-binding domain homology <ATN>
F:768-791/Domain: transmembrane #status predicted <TM5>
F:830-855/Domain: transmembrane #status predicted <TM6>
F:856-936/Domain: intracellular #status predicted <INT4>
F:937-955/Domain: transmembrane #status predicted <TM7>
F:956-1004/Domain: extracellular #status predicted <EXT>
F:357/Active site: Asp (aspartylphosphate intermediate) #status predicted
F:489/Binding site: Asp (Lys) #status predicted
F:698,702,707/Active site: Asp, Asp, Lys #status predicted

Query Match 43.0%; Score 46; DB 2; Length 1004;
Best Local Similarity 40.0%; Pred. No. 30;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 6 VFGLKQNWGSAISN 20
:||||| :| :|
DB 864 LFGLRKHWDGRAVND 878

RESULT 9
S76134
Hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S76134
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1034 <KAN>
A:Cross-references: EMBL:D90914; GB:AB001339; NID:g165347; PIDN:BAAL8393.1; PID:g165348
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: phosphoenolpyruvate carboxylase

Query Match 43.0%; Score 46; DB 2; Length 1034;
Best Local Similarity 50.0%; Pred. No. 31;
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGDS 16
| | | | | |
DB 2 NLAVPAFGLSTNWSGN 17

RESULT 10
A75466
2-demethylmenaquinone 2-C-methyltransferase (EC 2.1.1.-) DR0859 [similarity] - *Deinococcus*
C:Species: *Deinococcus* radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C:Accession: A75466
R:White, O.; Elsen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250; MUID:20036896
A:Accession: A75466
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <WHI>
A:Cross-references: GB:AE001940; GB:AE000513; NID:g6458577; PIDN:AAF10437.1; PID:g645
C:Genetics:
A:Gene: DR0859
A:Map position: 1
C:Keywords: methyltransferase

Query Match 42.1%; Score 45; DB 2; Length 160;
Best Local Similarity 46.7%; Pred. No. 6;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 4 VGVFGLKQNWGSAI 18
:||||: | | |
DB 79 LGVFGVENGWEGVII 93

RESULT 11
S60558
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-330-1 and
C:Species: human immunodeficiency virus type 1, HIV-1
A:Variety: isolate CI-330-1; isolate CI-330-2; isolate CI-330-4
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C:Accession: S60558; S60559; S60560
R:Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Franssen, K.; Motte, J
AIDS 8, 21-26, 1994
A:Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cot
A:Reference number: S60521; MUID:94280700
A:Accession: S60558
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-300 <JAN>
A:Cross-references: EMBL:X72062; NID:g468697; PIDN:CAA50943.1; PID:g468698
A:Experimental source: isolate CI-330-1
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
A:Accession: S60559
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-300 <JAN>
A:Cross-references: EMBL:X72063; NID:g468699; PIDN:CAA50944.1; PID:g468700
A:Experimental source: isolate CI-330-2
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
A:Accession: S60560
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-300 <JAF>
A:Cross-references: EMBL:X72065; NID:g468701; PIDN:CAA50946.1; PID:g468702
A:Experimental source: isolate CI-330-4
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C:Genetics:
A:Gene: env
C:Superfamily: type E retrovirus env polyprotein
C:Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match 42.1%; Score 45; DB 2; Length 300;
Best Local Similarity 47.1%; Pred. No. 12;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 TVGVFGLKQNWGSAIS 19
| | | | | |
DB 117 TSGLFNSTWNNGTAIT 133

RESULT 12

S60526
envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-330-3) (frag
C:Species: human immunodeficiency virus type 1, HIV-1
A:Variety: isolate CI-330-3
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999
C:Accession: S60526
R:Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J.;
AIDS 8, 21-26, 1994
A:Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cote d
A:Reference number: S60521; MUID:94280700
A:Accession: S60526
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-300 <JAN>
A:Cross-references: EMBL:X72064; NID:q468784; PIDN:CAA50945.1; PID:q468785
A:Experimental source: isolate CI-330-3
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
C:Genetics:
A:Gene: env
C:Superfamily: type E retrovirus env polyprotein
C:Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match 42.1%; Score 45; DB 2; Length 300;
Best Local Similarity 47.1%; Pred. No. 12;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 TVGVFGLKQNWGSAIS 19
:|||||
Db 117 TSGLFNSTWNGTALT 133
:|||||

RESULT 13
G85354
hypothetical protein AT4g30340 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: G85354
R:anonymous, T. 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: G85354
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <STO>
A:Cross-references: GB:NC_001268; NID:g7269934; PIDN:CAB81027.1; GSPDB:GN00140
A:Gene: AT4g30340
A:Map position: 4

Query Match 42.1%; Score 45; DB 2; Length 490;
Best Local Similarity 46.7%; Pred. No. 20;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 VGVFGLKQNWGSAI 18
:|||||
Db 412 IEIFGLKQGHASV 426
:|||||

RESULT 14
S54586
probable membrane protein YMR279c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR021.05c
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 05-Nov-1999
C:Accession: S54586
R:Pearson, D.; Bowman, S.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54582
A:Accession: S54586
A:Molecule type: DNA

A:Residues: 1-540 <PEA>
A:Cross-references: EMBL:Z49704; NID:q825540; PIDN:CAA89777.1; PID:q825545; GSPDB:GNO
A:Experimental source: strain AS972
C:Genetics:
A:Gene: MIPS:YMR279c
A:Map position: 13R
C:Superfamily: aminotriazole resistance protein YML116W
C:Keywords: transmembrane protein
F:109-125/Domain: transmembrane #status predicted <TM1>
F:139-155/Domain: transmembrane #status predicted <TM2>
F:174-190/Domain: transmembrane #status predicted <TM3>
F:233-249/Domain: transmembrane #status predicted <TM4>
F:268-284/Domain: transmembrane #status predicted <TM5>
F:298-314/Domain: transmembrane #status predicted <TM6>
F:335-351/Domain: transmembrane #status predicted <TM7>
F:377-393/Domain: transmembrane #status predicted <TM8>
F:400-416/Domain: transmembrane #status predicted <TM9>
F:435-451/Domain: transmembrane #status predicted <TM10>
F:503-519/Domain: transmembrane #status predicted <TM11>

Query Match 42.1%; Score 45; DB 2; Length 540;
Best Local Similarity 36.8%; Pred. No. 22;
Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAIS 19
:|||||
Db 256 NVPTNIHGLSMDWTGSALA 274
:|||||

RESULT 15
A56594
Na+/K+-exchanging ATPase (EC 3.6.1.37) alpha chain - cat flea
N:Alternate names: sodium pump alpha subunit
C:Species: ctenocephalides felis (cat flea)
C:Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 22-Jun-1999
C:Accession: A56594
R:Reeves, S.A.; Yamanaka, M.K.
Insect Biochem. Mol. Biol. 23, 809-814, 1993
A:Title: Cloning and sequence analysis of the alpha subunit of the cat flea sodium pu
A:Reference number: A56594; MUID:93386186
A:Accession: A56594
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1037 <REE>
A:Cross-references: GB:S66043; NID:q432541; PIDN:AAB28239.1; PID:q432542
A:Note: sequence extracted from NCBI backbone (NCBLN:137976, NCBIP:137977)
C:Superfamily: Na+/K+-transporting ATPase alpha chain: ATPase nucleotide-binding doma
C:Keywords: ATP; glycoprotein; hydrolase; phosphoprotein; potassium transport; sodium
F:601-797/Domain: ATPase nucleotide-binding domain homology <ATP>
F:390/Active site: Asp (aspartylphosphate intermediate) #status predicted
F:497/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:522/Binding site: ATP (lys) #status predicted

Query Match 41.1%; Score 44; DB 2; Length 1037;
Best Local Similarity 33.3%; Pred. No. 65;
Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 6 VFGLKQNWGSAISN 20
:|||||
Db 897 LFGIRKQWDSKAVND 911
:|||||

Search completed: March 14, 2002, 09:10:43
Job time: 359 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:10:43 ; Search time 56.96 Seconds
(without alignments)
24.072 Million cell updates/sec

Title: US-09-765-739A-4
Perfect score: 98
Sequence: 1 NPTVALYGLKQDWNGVSA 18
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68: *
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	89.8	278	2 JE0216	28k surface antige
2	61	62.2	280	2 JE0217	28k surface antige
3	57	58.2	276	2 JE0218	28k surface antige
4	55	56.1	284	2 I40882	major antigenic pr
5	50	51.0	286	2 JE0219	28k surface antige
6	48	49.0	439	2 E82426	phosphoglycerate t
7	47	48.0	341	2 S73685	hypothetical prote
8	46	46.9	133	2 JE0221	28k surface antige
9	44	44.9	1649	2 C86822	hypothetical prote
10	43.5	44.4	122	2 G64558	hypothetical prote
11	43.5	44.4	299	2 T23932	hypothetical prote
12	43	43.9	480	2 T15839	hypothetical prote
13	43	43.9	918	2 S04255	regulatory protein
14	42.5	43.4	328	2 G71838	probable nadh oxid
15	42.5	43.4	1465	2 S31282	TyB protein - years
16	42.5	43.4	1467	2 PC1253	TyB protein - years
17	42.5	43.4	1802	2 S52611	TyB protein - years
18	42.5	43.4	1803	2 S56894	TyB protein - years
19	42	42.9	298	2 S75205	hypothetical prote
20	42	42.9	389	2 T43979	hypothetical prote
21	42	42.9	412	2 T09313	immediate-early pr
22	42	42.9	656	1 S59631	endo-1,4-beta-xyla
23	42	42.9	661	1 S59633	endo-1,4-beta-xyla
24	42	42.9	691	2 B75622	hypothetical prote
25	42	42.9	1023	2 A47296	thiazide-sensitive
26	42	42.9	2894	2 C54474	hypothetical prote
27	41.5	42.3	122	2 A71950	hypothetical prote
28	41	41.8	243	2 A72669	probable glutaredo
29	41	41.8	290	2 S76787	hypothetical prote

30	41	41.8	373	2 G75073	hypothetical prote
31	41	41.8	412	2 T18760	hypothetical prote
32	41	41.8	459	2 T10307	hypothetical prote
33	41	41.8	544	2 S25101	CTP synthase (EC 6
34	41	41.8	585	2 B75265	probable acid--CoA
35	41	41.8	624	2 A83237	hypothetical prote
36	41	41.8	716	2 A75595	hypothetical prote
37	41	41.8	757	2 C84120	subtilisin-type pr
38	41	41.8	1232	2 T21018	hypothetical prote
39	40.5	41.3	89	2 E53226	retrovirus-related
40	40.5	41.3	358	2 B64427	carbamoyl-phosphat
41	40.5	41.3	982	2 S00954	pol polyprotein -
42	40.5	41.3	1356	2 F84486	probable retroelem
43	40	40.8	105	2 F83935	hypothetical prote
44	40	40.8	125	2 H83068	hypothetical prote
45	40	40.8	162	2 F82084	probable 2-demethy

ALIGNMENTS

RESULT 1
JE0216
28k surface antigen 3 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0216
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kda surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0216
A:Molecule type: DNA
A:Residues: 1-278 <RED>
A:Cross-references: GB:AF062761

Query Match 89.8%; Score 88; DB 2; Length 278;
Best Local Similarity 83.3%; Pred. No. 9.5e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NPTVALYGLKQDWNGVSA 18
Db 60 NPTVALYGLKQDWEGISS 77
|||||
|:|

RESULT 2
JE0217
28k surface antigen 4 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0217
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kda surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0217
A:Molecule type: DNA
A:Residues: 1-280 <RED>
A:Cross-references: GB:AF062761

Query Match 62.2%; Score 61; DB 2; Length 280;
Best Local Similarity 60.0%; Pred. NO. 0.023;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NPTVALYGLKQDWNG 15
Db 60 NTTICGVFLKQDWGD 74
| | : : ||||| : |

```

RESULT 3
JE0218
28k surface antigen 5 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0218
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0218
A:Molecule type: DNA
A:Residues: 1-276 <RED>
A:Cross-references: GB:AF062761

Query Match 58.2%; Score 57; DB 2; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.1;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNG 15
| | | | | | | | | |
Db 59 NTTVGVLKQNDG 73

RESULT 4
I40882
major antigenic protein - heartwater rickettsia
C:Species: Cowdria ruminantium (heartwater rickettsia)
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40882; S42827
R:van Vliet, A.H.; Jongejans, F.; van Kleef, M.; van der Zeijst, B.A.
Infect. Immun. 62, 1451-1456, 1994
A:Title: Molecular cloning, sequence analysis, and expression of the gene encoding the
A:Reference number: I40882; MUID:94178956
A:Accession: I40882
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-284 <RES>
A:Cross-references: EMBL:X74250; NID:g454266; PIDN:CAA52309.1; PID:g454267
C:Genetics:
A:Gene: map1

Query Match 56.1%; Score 55; DB 2; Length 284;
Best Local Similarity 64.3%; Pred. No. 0.22;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 TVALYGLKQDNGV 16
| | | | | | | | | |
Db 63 TRAVFGLKQDNGV 76

RESULT 5
JE0219
28k surface antigen 2 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0219
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0219
A:Molecule type: DNA
A:Residues: 1-286 <RED>
A:Cross-references: GB:AF062761

Query Match 51.0%; Score 50; DB 2; Length 286;
Best Local Similarity 50.0%; Pred. No. 1.5;

```

```

Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWN 14
| | | | | | | | | |
Db 60 NTTVGVEGIEQDWD 73

RESULT 6
E82426
phosphoglycerate transport system transcription regulator protein Pgta VCA0704 [impor
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: E82426
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
L., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: E82426
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-439 <HEI>
A:Cross-references: GB:AE004400; GB:AE003853; NID:g9658121; PIDN:AAF96603.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0704
A:Map position: 2

Query Match 49.0%; Score 48; DB 2; Length 439;
Best Local Similarity 50.0%; Pred. No. 4.8;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNGV 16
| | | | | | | | | |
Db 53 DPTQAMSHIQPDWNGV 68

RESULT 7
S73685
hypothetical protein yibD - Mycoplasma pneumoniae (strain ATCC 29342)
N:Alternate names: hypothetical protein P01_orf341
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C:Accession: S73685
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885
A:Accession: S73685
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-341 <HIM>
A:Cross-references: EMBL:AE000034; GB:U00089; NID:g1674031; PIDN:AAB96007.1; PID:g167
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Gene: yibD
A:Genetic code: SGC3

Query Match 48.0%; Score 47; DB 2; Length 341;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LKQDNGV 16
| | | | | | | | | |
Db 26 LKQDNGV 33

RESULT 8
JE0221

```

28k surface antigen 2 - Ehrlichia canis
 C:Species: Ehrlichia canis
 C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 21-Jul-2000
 C:Accession: JF0221
 R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R.
 Biochem. Biophys. Res. Commun. 247, 636-643, 1998
 A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
 A:Reference number: JE0216; MUID:98321180
 A:Accession: JE0221
 A:Molecule type: DNA
 A:Residues: 1-133 <RED>
 A:Cross-references: GB:AF062762; NID:g3327964; PIDN:AAC26722.1; PID:g3327966

Query Match 46.9%; Score 46; DB 2; Length 133;
 Best Local Similarity 53.8%; Pred. No. 3;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNG 15
 | : ||||| : |
 Db 62 TTVVYGLKENWAG 74

RESULT 9
 C86822
 hypothetical protein yqbK [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
 C:Species: Lactococcus lactis subsp. lactis
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
 C:Accession: C86822
 R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissbach, J.; Ehrlich
 Genome Res. in press, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium.
 A:Reference number: A86625
 A:Accession: C86822
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1649 <STO>
 A:Cross-references: GB:AE005176; NID:g12724583; PIDN:AAK05677.1; GSPDB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: yqbK

Query Match 44.9%; Score 44; DB 2; Length 1649;
 Best Local Similarity 41.2%; Pred. No. 84;
 Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVS 17
 : || : ||| |
 Db 1601 SPTGTGFKLNQNNSTS 1617

RESULT 10
 G64558
 hypothetical protein HP0311 - Helicobacter pylori (strain 26695)
 C:Species: Helicobacter pylori
 C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 C:Accession: G64558
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliak, H.G.; Glodek, A.; McKenne
 son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
 Nature 388, 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; MUID:97394467
 A:Accession: G64558
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-122 <TOM>
 A:Cross-references: GB:AE000549; GB:AE000511; NID:g2313403; PIDN:AAD07384.1; PID:g231341

Query Match 44.4%; Score 43.5; DB 2; Length 122;

Best Local Similarity 56.2%; Pred. No. 7;
 Matches 9; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 NPTVALYGLKQDWNG 15
 | : ||||| : |
 Db 55 NQILAFYGLKINDWQG 70

RESULT 11
 T23932
 hypothetical protein R05D7.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T23932
 R:Dobson, R.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19819
 A:Accession: T23932
 A>Status: preliminary; translated from GB/EMBL/DBBJ
 A:Molecule type: DNA
 A:Residues: 1-299 <WIL>
 A:Cross-references: EMBL:Z81105; PIDN:CAB03219.1; GSPDB:GN00019; CESP:R05D7.4
 A:Experimental source: clone R05D7
 C:Genetics:
 A:Gene: CESP:R05D7.4
 A:Map position: 1
 A:Introns: 29/3; 122/1; 162/1; 231/2; 268/3
 C:Superfamily: tropinesterase

Query Match 44.4%; Score 43.5; DB 2; Length 299;
 Best Local Similarity 47.4%; Pred. No. 18;
 Matches 9; Conservative 4; Mismatches 3; Indels 3; Gaps 1;

QY 1 NPTVALYGL---KQDWNGV 16
 : | : ||| : ||| |
 Db 46 SPLVIVHGLFGKQNNNSV 64

RESULT 12
 T15839
 hypothetical protein C54G7.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 15-Sep-2000
 C:Accession: T15839
 R:Du, Z.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid C54G7.
 A:Reference number: Z18416
 A:Accession: T15839
 A>Status: preliminary; translated from GB/EMBL/DBBJ
 A:Molecule type: DNA
 A:Residues: 1-480 <DUZ>
 A:Cross-references: EMBL:U40410; NID:g1065453; PID:g1065454; PIDN:AAA81391.1; CESP:CS
 C:Genetics:
 A:Gene: CESP:C54G7.2
 A:Introns: 27/3; 76/2; 102/2; 185/3; 283/1; 384/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein C08F8.4

Query Match 43.9%; Score 43; DB 2; Length 480;
 Best Local Similarity 60.0%; Pred. No. 34;
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNGVS 17
 ||| | | | | |
 Db 351 TVATYTLISAVWHGVS 365

RESULT 13
 S04255
 regulatory protein qa-1S - Neurospora crassa
 C:Species: Neurospora crassa

C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 21-Jul-2000
C;Accession: S04255; A23941; E31277
R;Geever, R.F.; Huiet, L.; Baum, J.A.; Tyler, B.M.; Rutledge, B.J.; Case, M.
J. Mol. Biol. 207, 15-34, 1989
A;Title: DNA sequence, organization and regulation of the qa gene cluster of Neurospora
A;Reference number: S04250; MUID:89293848
A;Accession: S04255
A;Molecule type: DNA
A;Residues: 1-918 <GIL>
A;Cross-references: EMBL:X14603; NID:g3060; PIDN:CAA32753.1; PID:g295929
R;Huiet, L.; Giles, N.H. Proc. Natl. Acad. Sci. U.S.A. 83, 3381-3385, 1986
A;Title: The qa repressor gene of Neurospora crassa: wild-type and mutant nucleotide sequence
A;Reference number: A23941; MUID:86205901
A;Accession: A23941
A;Molecule type: DNA
A;Residues: 1-904, 'R', 906, 'R', 908-918 <GIL2>
A;Cross-references: GB:M13208; NID:g169869; PIDN:AAA33612.1; PID:g169870
A;Note: this sequence has been revised in reference S04250
C;Genetics:
A;Gene: qa-1S
A;Introns: 76/1
C;Superfamily: shikimate dehydrogenase homology
C;Keywords: DNA binding; repressor; transcription regulation
F;649-877/Domain: shikimate dehydrogenase homology <SKD>

Query Match 43.9%; Score 43; DB 2; Length 918;
Best Local Similarity 57.1%; Pred. No. 67;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 5 ALYGLKQNGVSA 18
|||||
Db 684 ALYGLKQNGVSA 18

RESULT 14

G71838
Probable nadh oxidoreductase I - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: Strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C;Accession: G71838
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A;Reference number: A71800; MUID:99120557
A;Accession: G71838
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-328 <ARN>
A;Cross-references: GB:AE001545; GB:AE001439; NID:g4155776; PIDN:AAD06772.1; PID:g415579
A;Experimental source: strain J99
C;Genetics:
A;Gene: nuoF

Query Match 43.4%; Score 42.5; DB 2; Length 328;
Best Local Similarity 64.3%; Pred. No. 28;
Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 NPTVALYGLK-QDW 13
|||||
Db 203 NPTVALYGLK-QDW 216

RESULT 15

S31262
TyB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty4 (fragment)
C;Species: Saccharomyces cerevisiae
C;Date: 18-Jun-1993 #sequence_revision 25-Apr-1997 #text_change 26-May-2000
C;Accession: S31262

R;Janetzky, B.; Lehle, L.
J. Biol. Chem. 267, 19798-19805, 1992
A;Title: Ty4, a new retrotransposon from Saccharomyces cerevisiae, flanked by tau-ele
A;Reference number: S31261; MUID:93015829
A;Accession: S31262
A;Molecule type: DNA
A;Residues: 1-1465 <JAN>
A;Cross-references: EMBL:S46865
C;Genetics:
A;Mobile element: retrotransposon Ty4
C;Superfamily: TyB protein

Query Match 43.4%; Score 42.5; DB 2; Length 1465;
Best Local Similarity 69.2%; Pred. No. 1.3e+02;
Matches 9; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

Qy 5 ALYGLKQ---DWN 14
|||||
Db 1077 ALYGLKQSPKEWN 1089

Search completed: March 14, 2002, 09:10:44
Job time: 360 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:24:49 ; Search time 32.82 seconds
(without alignments)
22.343 Million cell updates/sec

Title: US-09-765-739A-3

Perfect score: 107

Sequence: 1 NTTGVFGLKQNWGSAISN 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	DB ID	Description
1	46	43.0	1004	1 ATNA_ARTSF	P28774 artemia san
2	46	43.0	1034	1 CAPP_SYNY3	P47299 synechocyst
3	45	42.1	540	1 YMBM_YEAST	Q03263 saccharomyc
4	43	40.2	86	1 GYPM_HALME	Q02238 halobacteri
5	43	40.2	422	1 YF02_MYCPN	P75285 mycoplasma
6	43	40.2	428	1 YB01_MYCPN	P75568 mycoplasma
7	43	40.2	438	1 YC05_MYCPN	P75571 mycoplasma
8	43	40.2	648	1 VP3_BPPH6	P11129 bacterioph
9	42.5	39.7	1538	1 GLSF_PORPU	P51375 porphyra pu
10	42	39.3	347	1 UTR2_YEAST	P32623 saccharomyc
11	42	39.3	1041	1 ATNA_DROME	P13607 drosophila
12	41	38.3	173	1 YYAR_BACSU	P37506 bacillus su
13	41	38.3	233	1 RR2_CYAPA	P48132 cyanophora
14	41	38.3	282	1 BLOB_HELPJ	Q29288 hellobacte
15	41	38.3	282	1 BLOB_HELPJ	Q29288 hellobacte
16	41	38.3	330	1 Y355_SYNY3	P74436 synechocyst
17	41	38.3	377	1 YA67_METH	Q27139 methanobact
18	41	38.3	400	1 OAS1_HUMAN	P00973 homo sapien
19	41	38.3	426	1 BZTB_RHOCA	Q32664 rhodobacter
20	41	38.3	756	1 K6PF_CABEL	Q27483 caenorhabdi
21	40.5	37.9	303	1 NORF_ALTEA	Q91CJ3 alteromonas
22	40	37.4	226	1 SPCL_CANFA	Q28250 canis famil
23	40	37.4	226	1 SPCL_CANFA	Q28250 canis famil
24	40	37.4	354	1 DHAS_METJA	Q57658 methanococc
25	40	37.4	465	1 YHJA_ECOLI	P37197 escherichia
26	40	37.4	473	1 XYLA_CLOSR	P48790 clostridium
27	40	37.4	543	1 CH60_MYCPN	P78012 mycoplasma
28	40	37.4	607	1 HRA1_XANCV	P80151 xanthomonas
29	40	37.4	689	1 SYGB_PASNU	P57905 pasteurella
30	40	37.4	942	1 ENV_CAEVG	P31627 caprine art
31	39.5	36.9	303	1 NQRE_ALTHA	Q91CJ4 alteromonas
32	39.5	36.9	303	1 NORF_PHOPO	Q91CJ1 photobacter
33	39.5	36.9	830	1 FAR1_YEAST	P21268 saccharomyc

34	39	36.4	161	1 HCRC_THAAR	O33818 thauera aro
35	39	36.4	231	1 IFEL_CABEL	O45551 caenorhabdi
36	39	36.4	286	1 TVSY_BPT4	P00471 bacterioph
37	39	36.4	308	1 PQOB_KLEPN	P27504 klebsiella
38	39	36.4	326	1 GBLP_NICPL	P93340 nicotiana p
39	39	36.4	326	1 GBLP_TOBAC	P49026 nicotiana t
40	39	36.4	371	1 Y028_ARCFU	O30207 archaeoglob
41	39	36.4	388	1 P2X4_HUMAN	Q09571 homo sapien
42	39	36.4	527	1 LAC5_TRAVE	Q12717 trametres ve
43	39	36.4	527	1 LAC5_TRAVI	O99056 trametres vi
44	39	36.4	1031	1 ATNA_HYDAT	P55317 hydra atten
45	39	36.4	1033	1 S190_YEAST	P36123 saccharomyc

ALIGNMENTS

RESULT	1
ATNA_ARTSF	
ID	ATNA_ARTSF
AC	P28774;
DT	01-DEC-1992 (Rel. 24, Created)
DT	01-DEC-1992 (Rel. 24, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	SODIUM/POTASSIUM-TRANSPORTING ATPASE ALPHA CHAIN (EC 3.6.3.9) (SODIUM PUMP) (NA+/K+ ATPASE).
OS	Artemia sanctificans (Brine shrimp) (Artemia franciscana).
OC	Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda; Anostraca;
OC	Artemiidae; Artemia.
OX	NCBI_TaxID:6661;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=92039032; PubMed=1657719;
RA	Macias M.T., Martinez J.L., Palmero I., Sastre L.;
RT	"Cloning of a cDNA encoding an Artemia franciscana Na/K ATPase alpha-
RT	subunit.";
RL	Gene 105:197-204(1991).
CC	FUNCTION: THIS IS THE CATALYTIC COMPONENT OF THE ACTIVE ENZYME, WHICH CATALYZES THE HYDROLYSIS OF ATP COUPLED WITH THE EXCHANGE OF NA & K IONS ACROSS THE PLASMA MEMBRANE. THIS ACTION CREATES THE ELECTROCHEMICAL GRADIENT OF NA & K, PROVIDING THE ENERGY FOR ACTIVE TRANSPORT OF VARIOUS NUTRIENTS.
CC	- - CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(IN) + NA(+)(OUT) = ADP + PHOSPHATE + H(+)(OUT) + NA(+)(IN).
CC	- - SUBUNIT: COMPOSED OF THREE SUBUNITS: ALPHA (CATALYTIC), BETA AND GAMMA.
CC	- - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC	- - SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2 ATPASES).
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; X56650; CAA39972.1; -
DR	PIR; JH0470; JH0470.
DR	InterPro; IPR001757; E1-E2_ATPase.
DR	InterPro; IPR001454; Hydrolase.
DR	InterPro; IPR000661; Na_H_K_ATPase.
DR	Pfam; PF00122; E1-E2_ATPase; 1.
DR	Pfam; PF00702; Hydrolase; 1.
DR	Pfam; PF00689; Na_K_ATPase_C; 1.
DR	Pfam; PF00690; Na_K_ATPase_N; 1.
DR	PRINTS; PR00119; CATATPASE.
DR	PRINTS; PR00121; NAKATPASE.
DR	PROSITE; PS00154; ATPASE_E1_E2; 1.
KW	Hydrolase; Sodium/potassium transport; Transmembrane;
KW	Phosphorylation; ATP-binding.
FT	TRANSMEM 76 BY SIMILARITY. 96

FT TRANSMEM 110 126 BY SIMILARITY.
 FT TRANSMEM 272 294 BY SIMILARITY.
 FT TRANSMEM 301 329 BY SIMILARITY.
 FT TRANSMEM 768 791 BY SIMILARITY.
 FT TRANSMEM 828 855 BY SIMILARITY.
 FT TRANSMEM 897 918 BY SIMILARITY.
 FT TRANSMEM 934 959 BY SIMILARITY.
 FT MOD_RES 357 357 PHOSPHORYLATION (PROBABLE).
 FT BINDING 489 489 ATP (BY SIMILARITY).
 SQ SEQUENCE 1004 AA; 110699 MW; CE4E6BECE19A78C7 CRC64;

Query Match 43.0%; Score 46; DB 1; Length 1004;
 Best Local Similarity 40.0%; Pred. No. 11;
 Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 6 VFGLKQWDSALSN 20
 :|||::|||:::
 Db 864 LFGLRKHWSRAVND 878

RESULT 2
 CAPP_SYNY3 STANDARD; PRT; 1034 AA.
 AC P74299;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PHOSPHOENOLPYRUVATE CARBOXYLASE (EC 4.1.1.31) (PEPCASE) (PEPC).
 GN PPC OR SL0920.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirasawa M., Suglura M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PCC6803. II. Sequence determination of the
 entire genome and assignment of potential protein-coding regions.;"
 RL DNA Res. 3:109-136(1996).
 CC -!- FUNCTION: TO FORM OXALOACETATE, A FOUR-CARBON DICARBOXYLIC ACID
 CC SOURCE FOR THE TRICARBOXYLIC ACID CYCLE (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ORTHOPHOSPHATE + OXALOACETATE -> H(2)O +
 CC PHOSPHOENOLPYRUVATE + CO(2).
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 CC -!- SIMILARITY: BELONGS TO THE PEPCASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D90914; BAA18393.1;
 CC InterPro: IPR001449; PEPCase.
 CC Pfam: PF00311; PEPCase_3.
 CC PRINTS: PR00150; PEPCARXLA2.
 CC PROSITE: PS00393; PEPCASE_2; 1.
 CC PROSITE: PS00781; PEPCASE_1; 1.
 CC Lyase: Carbon dioxide fixation; Allosteric enzyme;
 KW Tricarboxylic acid cycle; Complete proteome.
 FT ACT_SITE 203 203 BY SIMILARITY.
 FT ACT_SITE 680 680 BY SIMILARITY.
 SQ SEQUENCE 1034 AA; 118940 MW; 76DF6061BAB7235D CRC64;

Query Match 43.0%; Score 46; DB 1; Length 1034;
 Best Local Similarity 50.0%; Pred. No. 11;
 Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 NTTVGVFGLKQWDSGS 16
 :|::|||:::
 Db 2 NLAVPAFGLSTNWSGN 17

RESULT 3
 YM8M_YEAST STANDARD; PRT; 540 AA.
 AC Q03263;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE HYPOTHETICAL 59.6 KDA PROTEIN IN DSK2-CAT8 INTERGENIC REGION.
 GN YMR279C OR YMR021.05C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
 CC AS THE DRUG RESISTANCE TRANSLUCASE FAMILY). DHA12 SUBFAMILY.
 CC STRONG, TO YEAST ATR1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z49704; CAA89777.1;
 CC SGD: S0004892; YMR279C.
 KW Hypothetical protein; Transport; Transmembrane.
 FT TRANSMEM 62 82 POTENTIAL.
 FT TRANSMEM 109 129 POTENTIAL.
 FT TRANSMEM 132 152 POTENTIAL.
 FT TRANSMEM 170 190 POTENTIAL.
 FT TRANSMEM 204 224 POTENTIAL.
 FT TRANSMEM 233 253 POTENTIAL.
 FT TRANSMEM 273 293 POTENTIAL.
 FT TRANSMEM 296 316 POTENTIAL.
 FT TRANSMEM 335 355 POTENTIAL.
 FT TRANSMEM 373 393 POTENTIAL.
 FT TRANSMEM 399 419 POTENTIAL.
 FT TRANSMEM 430 450 POTENTIAL.
 FT TRANSMEM 462 482 POTENTIAL.
 FT TRANSMEM 503 523 POTENTIAL.
 SQ SEQUENCE 540 AA; 59561 MW; 687D06CB0D70AF91 CRC64;

Query Match 42.1%; Score 45; DB 1; Length 540;
 Best Local Similarity 36.8%; Pred. No. 8.4;
 Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 NTTVGVFGLKQWDSAYS 19
 :|::|||::|||:
 Db 256 NVPTNIHGLSLMDWTGSALA 274

RESULT 4
 GVPM_HALME STANDARD; PRT; 86 AA.
 ID GVPM_HALME
 AC Q02238;

DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE GVPN PROTEIN.
GN GVPN.
OS Halobacterium mediterranei (Haloferax mediterranei).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloferax.
OX NCBI_TaxID=2252;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1411;
RA MEDLINE=93021102; PubMed=1404376;
RA Engler C., Krueger K., Offner S., Pfeifer F.;
RT "Three different but related gene clusters encoding gas vesicles in
RT halophilic archaea";
RL J. Mol. Biol. 227:586-592(1992).
CC -!- FUNCTION: COULD BE IMPORTANT FOR THE SHAPE DETERMINATION OF THE
CC GAS VESICLE.
CC -!- SIMILARITY: HIGH, TO GVPJ AND TO GAS VESICLE PROTEINS TYPE A.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64701; CAA45955.1; -.
DR PIR; S28126; S28126.
DR InterPro; IPR000638; Gas_vesicle.
DR Pfam; PF00741; Gas_vesicle; 1.
DR ProDom; PD003598; Gas_vesicle; 1.
DR ProSITE; PS00234; GAS_VESICLE_A_1; 1.
DR ProSITE; PS00669; GAS_VESICLE_A_2; 1.
KW Gas vesicle.
SQ SEQUENCE 86 AA; 9355 MW; AB522F20C6F3666 CRC64;

Query Match 40.2%; Score 43; DB 1; Length 86;
Best Local Similarity 40.0%; Pred. No. 2.6;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 TTVGVFGLKQNDGS 16
II : : : : :
DB 51 TTMTAYGFENWDAT 65

RESULT 5
YF02_MYCPN STANDARD; PRT; 422 AA.
ID YF02_MYCPN STANDARD; PRT; 422 AA.
AC P75285;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MPN502 (P02_ORF422V).
GN MPN502 OR MP341.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RA MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SIMILARITY: BELONGS TO THE ADHESIN P1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000006; AAB95701.1; -.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 428 AA; 46908 MW; 991143BEF683A261 CRC64;

Query Match 40.2%; Score 43; DB 1; Length 428;
Best Local Similarity 70.0%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTVGVFGLK 10
III : : : :
DB 141 NTTIGAYGLK 150

RESULT 7
YC05_MYCPN STANDARD; PRT; 438 AA.
ID YC05_MYCPN STANDARD; PRT; 438 AA.
AC P75571;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000032; AAB95988.1; -.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 422 AA; 46141 MW; 03BE7D381C7C298 CRC64;

Query Match 40.2%; Score 43; DB 1; Length 422;
Best Local Similarity 70.0%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTVGVFGLK 10
III : : : :
DB 141 NTTIGAYGLK 150

RESULT 6
YB01_MYCPN STANDARD; PRT; 428 AA.
ID YB01_MYCPN STANDARD; PRT; 428 AA.
AC P75568;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MPN101 (C09_ORF428V).
GN MPN101 OR MP053.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RA MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SIMILARITY: BELONGS TO THE ADHESIN P1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000006; AAB95701.1; -.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 428 AA; 46908 MW; 991143BEF683A261 CRC64;

Query Match 40.2%; Score 43; DB 1; Length 428;
Best Local Similarity 70.0%; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTVGVFGLK 10
III : : : :
DB 141 NTTIGAYGLK 150

RESULT 7
YC05_MYCPN STANDARD; PRT; 438 AA.
ID YC05_MYCPN STANDARD; PRT; 438 AA.
AC P75571;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)

20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPERRETICAL PROTEIN MPN205 (GT9-ORF438V).
GN MPN205 OR MP626.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29242 / M129; PubMed=8948633;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelsbach R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Hermann R.
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: BELONGS TO THE ADHESIN P1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000060; AAB96274.1;
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 438 AA; 47925 MW; DQCB89515576AD7E CRC64;
SQ

Query Match 40.28; Score 43; DB 1; Length 438;
Best Local Similarity 70.08; Pred. No. 14;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTVGVGFLK 10
DB 140 NTVGVGFLK 149
|||||

RESULT 8
ID VP3_BPPH6 STANDARD; PRT; 648 AA.
AC P11129;
DT 01-JUL-1989 (Rel. 11, Created)
DE 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last annotation update)
DE P3. PROTEIN.
GN P3.
OS Bacteriophage phi-6.
OC Viruses; dsRNA viruses; Cystoviridae; Cystovirus.
OX NCBI_TaxID=10879;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-13.
RX MEDLINE=88160044; PubMed=3347997;
RA Gottlieb P., Metzger S., Romantschuk M., Carton J., Strassman J.,
RA Bamford D.H., Kalkinen N., Mindich L.;
RT Nucleotide sequence of the middle dsRNA segment of bacteriophage phi
RL Virology 163:183-190(1988).
CC -1- FUNCTION: P3 PROTEIN IS NECESSARY FOR ADSORPTION ONTO HOST
CC CELLS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M17462; AAA68485.1;
DR PIR; C28646; P3BPF6.
SQ

KW Envelope protein.
SQ SEQUENCE 648 AA; 69178 MW; B188DFE02ACC54E3 CRC64;

Query Match 40.28; Score 43; DB 1; Length 648;
Best Local Similarity 41.2%; Pred. No. 22;
Matches 7; Conservative 7; Mismatches 1; Indels 2; Gaps 1;

QY 6 VFC--LKQNDGSAISN 20
DB 181 IFGWYKMDWEGSAVAD 197
|||||

RESULT 9
ID GLSF_PORPU STANDARD; PRT; 1538 AA.
AC P51375;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE FERREDOXIN-DEPENDENT GLUTAMATE SYNTHASE (EC 1.4.7.1) (FD-GOGAT).
GN GLTB.
OS Porphyra purpurea.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AVONPORT.
RA Reith M.E., Munnolland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC -1- CATALYTIC ACTIVITY: 2 L-GLUTAMATE + 2 OXIDIZED FERREDOXIN =
CC L-GLUTAMINE + 2-OXOGLUTARATE + 2 REDUCED FERREDOXIN.
CC -1- COFACTOR: BINDS A 3FE-4S CLUSTER; FAD AND FMN.
CC -1- PATHWAY: GLUTAMINE SYNTHETASE/GOGAT PATHWAY WHICH IS INVOLVED
CC IN THE ASSIMILATION OF AMMONIA.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST STROMA.
CC -1- SIMILARITY: TO OTHER GLUTAMATE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U38804; AAC08261.1;
DR InterPro; IPR002489; DUF14.
DR InterPro; IPR003009; FMN_enzyme.
DR InterPro; IPR002932; Glu_synthase.
DR Pfam; PF01493; DUF14; 1.
DR Pfam; PF01645; Glu_synthase; 1.
KW Oxidoreductase; Iron-sulfur; 3Fe-4S; Flavoprotein; FAD; FMN;
KW Chloroplast; Glutamate biosynthesis.
RN DOMAIN 34 384
FT METAL 1162 1162 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
FT METAL 1168 1168 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
FT METAL 1173 1173 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
FT NP_BIND 1109 1166 FMN (BY SIMILARITY).
SQ SEQUENCE 1538 AA; 168710 MW; 3F647CDB2F5C77CF CRC64;

Query Match 39.7%; Score 42.5; DB 1; Length 1538;
Best Local Similarity 39.1%; Pred. No. 64;
Matches 9; Conservative 4; Mismatches 5; Indels 5; Gaps 1;

QY 1 NTVGVGVF-----GLKQNDGSAI 18
DB 345 NTEISDFEYYSGLQEPWDGPAL 367
|||||

```

RESULT 10
UTR2_YEAST STANDARD; PRT; 347 AA.
ID UTR2_YEAST
AC P32623;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE UTR2 PROTEIN (UNKNOWN TRANSCRIPT 2 PROTEIN).
GN UTR2 OR YEL040W OR SYGP-ORF18
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Mulligan J.T., Dietrich F.S., Hennessey K.M., Sehl P., Komp C.,
RA Wei Y., Taylor P., Nakahara K., Roberts D., Davis R.W.;
RA Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Bero A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yellon M., Botstein D., Davis R.W.;
RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=B-641;
RX MEDLINE=94016558; PubMed=8411151;
RA Melnick L., Sherman F.;
RT "The gene clusters ARC and COR on chromosomes 5 and 10, respectively,
RT of Saccharomyces cerevisiae share a common ancestry.";
RJ J. Mol. Biol. 233:372-388(1993).
CC 1- SIMILARITY: SOME, TO YEAST YGR189C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U18779; AAB65002.1; ALT_INIT.
DR EMBL; L22173; AAA34941.1; -.
DR EMBL; S65964; AAD13975.1; -.
DR EMBL; S66130; AAB28444.1; -.
DR PIR; S30839; S30839.
DR HSP; P23904; IAJ0.
DR SGD; S0000766; UTR2.
DR InterPro; IPR000757; Glyco_hydro_16.
DR Pfam; PF00722; Glyco_hydro_16; 1.
FT DOMAIN 234 322 SER-RICH.
FT DOMAIN 269 283 POLY-SER.
FT CONFLICT 10 10 L -> V (IN REF. 3).
FT CONFLICT 171 171 A -> R (IN REF. 3).
FT CONFLICT 234 234 S -> C (IN REF. 3).
SQ SEQUENCE 347 AA; 36592 MW; 1E8AFB862C4BB328 CRC64;

Query Match 39.3%; Score 42; DB 1; Length 347;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 TVGVFGLKQNWGDSAIN 20
I: | | | | |
DB 148 TIWASGGEINWDASDISN 165

```

```

RESULT 11
ATNA_DROME STANDARD; PRT; 1041 AA.
ID ATNA_DROME
AC P13607; O61494; Q9VDG6; Q9VDG7; Q9VDG8;
DT 01-JAN-1990 (Rel. 13, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SODIUM/POTASSIUM-TRANSPORTING ATPASE ALPHA CHAIN (EC 3.6.3.9) (SODIUM
DE PUMP) (NA+/K+ ATPASE).
GN NA-P OR ATP-ALPHA OR CG5670.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eurygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=CANTON-S, AND OREGON-R;
RX MEDLINE=89231618; PubMed=2540956;
RA Lebovitz R.M., Takeyasu K., Fambrough D.M.;
RT "Molecular characterization and expression of the (Na+ + K+)-ATPase
RT alpha-subunit in Drosophila melanogaster.";
RL EMBO J. 8:193-202(1989).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Head;
RX MEDLINE=98311194; PubMed=9648860;
RA Sun B., Wang W., Salvaterra P.M.;
RT "Functional analysis and tissue-specific expression of Drosophila
RT Na+,K+-ATPase subunits.";
RJ J. Neurochem. 71:142-151(1998).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

```



```

Db 74 NNCIGRIKIRSNWNGYAL 91
RESULT 13
RR2_CYAPA STANDARD; PRT; 233 AA.
AC P48132;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYANELLE 30S RIBOSOMAL PROTEIN S2.
GN RS2.
OS Cyanophora paradoxa.
OC Cyanelle.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LB555 / PRINGSHEIM;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RL Plant Mol. Biol. Rep. 13:327-332(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LB555 / PRINGSHEIM;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
RT "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RT the genetic complexity of a primitive plastid.";
RL (In) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL Schwemmler W. (eds.);
RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL (1997).
CC -1- SIMILARITY: BELONGS TO THE S2P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U30821; AAA81258.1; -.
CC Mendel: 7923; CYAPA: rps2.1.
CC InterPro: IPR001865; Ribosomal_S2.
CC Pfam: PF00318; Ribosomal_S2; 1.
CC PRINTS: PR00395; RIBOSOMALS2.
CC PROSITE: PS00962; RIBOSOMAL_S2_1; 1.
CC PROSITE: PS00963; RIBOSOMAL_S2_2; 1.
CC Ribosomal protein; Cyanelle.
KW SEQUENCE 233 AA; 26198 MW; 06D3BAD7DA9A883F CRC64;
SQ
Query Match 38.3%; Score 41; DB 1; Length 233;
Best Local Similarity 36.8%; Pred. No. 16;
Matches 7; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
Qy 2 TTVGVFLKQKQWDSATSN 20
I I I I I I I I
Db 85 TRCGAFVINQRWLGGTLTN 103
RESULT 14
BIOB_HELPJ STANDARD; PRT; 282 AA.
ID BIOB_HELPJ
AC Q9ZJK8;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BIOTIN SYNTHASE (EC 2.8.1.6) (BIOTIN SYNTHETASE).
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,

```

RA Venter J.C.:
RT "The complete genome sequence of the gastric pathogen Helicobacter
RL pylori.";
RL Nature 388:539-547(1997).
CC -!- CATALYTIC ACTIVITY: DETHIOBIOTIN + (S) = BIOTIN.
CC -!- PATHWAY: LAST STEP IN BIOTIN BIOSYNTHESIS PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000640; AAD08448.1; -.
DR TIGR; HP1406; -.
DR InterPro; IPR002684; Biotin_synth.
DR Pfam; PF01792; Biotin_synth; 1.
KW Biotin biosynthesis; Iron-sulfur; Transferase; Complete proteome.
FT METAL 17 17
FT METAL 21 21 IRON-SULFUR (POTENTIAL).
FT METAL 24 24 IRON-SULFUR (POTENTIAL).
SQ SEQUENCE 282 AA; 31475 MW; 8EAD8B801DF151EE CRC64;

Query Match 38.3%; Score 41; DB 1; Length 282;
Best Local Similarity 50.0%; Pred. NO. 19;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Qy 5 GVFGKONWD 14
|:|:|:|:|:
Db 157 GIGGLNESWE 166

Search completed: March 14, 2002, 09:24:50
Job time: 906 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:24:09 ; Search time 103.37 Seconds
(without alignments)
28.301 Million cell updates/sec

Title: US-09-765-739A-3
Perfect score: 107
Sequence: 1 NTTGVFGLKQNWGSAISN 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mhc:*
- 7: sp_mmmal:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	100.0	276	2 085358	085358 ehrlichia c
2	107	100.0	281	2 09AC19	09AC19 ehrlichia c
3	91	85.0	280	2 085357	085357 ehrlichia c
4	88	82.2	245	2 09RH35	09RH35 ehrlichia c
5	88	82.2	276	2 085817	085817 ehrlichia c
6	88	82.2	280	2 085816	085816 ehrlichia c
7	88	82.2	280	2 092GM9	092GM9 ehrlichia c
8	88	82.2	280	2 052107	052107 ehrlichia c
9	82	76.6	288	2 092GJ2	092GJ2 ehrlichia c
10	80	74.8	280	2 09F473	09F473 ehrlichia c
11	80	74.8	280	2 09ADV3	09ADV3 ehrlichia c
12	73	68.2	286	2 052105	052105 ehrlichia c
13	72	67.3	278	2 09R8A9	09R8A9 ehrlichia c
14	72	67.3	278	2 09R8A8	09R8A8 ehrlichia c
15	72	67.3	278	2 09R8A7	09R8A7 ehrlichia c
16	72	67.3	278	2 09R8A6	09R8A6 ehrlichia c
17	72	67.3	278	2 09R8A5	09R8A5 ehrlichia c
18	72	67.3	278	2 09R3J3	09R3J3 ehrlichia c
19	72	67.3	278	2 09F472	09F472 ehrlichia c

20	72	67.3	307	2 09ZGJ1	09ZGJ1 ehrlichia c
21	68	63.6	276	2 09F475	09F475 ehrlichia c
22	61.5	57.5	133	2 085360	085360 ehrlichia c
23	61.5	57.5	283	2 09F474	09F474 ehrlichia c
24	60	56.1	278	2 052106	052106 ehrlichia c
25	60	56.1	280	2 052104	052104 ehrlichia c
26	53	49.5	272	2 09AMF6	09AMF6 cowdria sp.
27	53	49.5	284	2 046327	046327 cowdria rum
28	53	49.5	284	2 09AFAL	09AFAL cowdria rum
29	53	49.5	737	13 090422	090422 brachydanio
30	52	48.6	290	2 09AEU3	09AEU3 cowdria rum
31	51	47.7	265	2 09AF99	09AF99 cowdria rum
32	51	47.7	270	2 09AF98	09AF98 cowdria rum
33	51	47.7	281	2 046328	046328 cowdria rum
34	51	47.7	281	2 09S6H1	09S6H1 cowdria rum
35	51	47.7	281	2 09S6H0	09S6H0 cowdria rum
36	51	47.7	287	2 046329	046329 cowdria rum
37	51	47.7	287	2 046331	046331 cowdria rum
38	51	47.7	287	2 09R425	09R425 cowdria rum
39	51	47.7	290	2 046324	046324 cowdria rum
40	51	47.7	290	2 046330	046330 cowdria rum
41	51	47.7	290	2 046332	046332 cowdria rum
42	51	47.7	290	2 046333	046333 cowdria rum
43	48	44.9	308	5 09W3R8	09W3R8 drosophila
44	47	43.9	1029	2 09A9L0	09A9L0 caulobacter
45	46	43.0	299	5 045707	045707 caenorhabdi

ALIGNMENTS

RESULT 1	
085358	
ID 085358	PRELIMINARY; PRT; 276 AA.
AC 085358	
DT 01-NOV-1998 (TrEMBLrel. 08, Created)	
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)	
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)	
DE 28 KDA MAJOR SURFACE ANTIGEN-5 (FRAGMENT).	
OS Ehrlichia chaffeensis.	
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;	
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.	
OX NCBI_TaxID=945;	
RN [1]	SEQUENCE FROM N.A.
RP STRAIN=ARKANSAS;	
RC STRAIN=ARKANSAS;	
RX MEDLINE=98321180; PubMed=9647746;	
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., BurrIDGE M.J.,	
RA Alleman A.R.;	
RT "Molecular characterization of a 28 kDa surface antigen gene family of the tribe Ehrlichiae."	
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).	
DR EMBL; AF062761; AAC26716.1; -.	
DR InterPro; IPR002566; Surface_Ag_msp4.	
DR Pfam; PF01617; Surface_Ag_2; 1.	
FT NON_TER 276	
SQ SEQUENCE 276 AA; 29782 MW; A9AB2A92263CA4EA CRC64;	
Query Match 100.0%; Score 107; DB 2; Length 276;	
Best Local Similarity 100.0%; Pred. No. 1.4e-09;	
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy 1 NTTGVFGLKQNWGSAISN 20	
Db 59 NTTGVFGLKQNWGSAISN 78	
RESULT 2	
09AC19	
ID 09AC19	PRELIMINARY; PRT; 281 AA.
AC 09AC19	
DT 01-JUN-2001 (TrEMBLrel. 17, Created)	

```
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN P28.
GN P28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL: U72291; AAC28673.1; -
SQ SEQUENCE 281 AA; 30343 MW; A995E5F7C4459AA9A CRC64;

Query Match 100.0%; Score 107; DB 2; Length 281;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQKQWDGSAISN 20
   |||||
DB 59 NTTGVFGLKQKQWDGSAISN 78
   |||||

RESULT 3
ID 085357 PRELIMINARY; PRT; 280 AA.
AC 085357;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 28 KDA MAJOR SURFACE ANTIGEN-4.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Allaman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL: AF062761; AAC26720.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30743 MW; FBB841DAF08EE4DC CRC64;

Query Match 85.0%; Score 91; DB 2; Length 280;
Best Local Similarity 84.2%; Pred. No. 6.1e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQKQWDGSAIS 19
   |||||
DB 60 NTTGVFGLKQKQWDGSAIS 78
   |||||

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN P28.
GN P28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=ST. VINCENT;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
DR EMBL: AF077735; AAC31548.1; -
DR InterPro: IPR001702; Gram_neg_porin.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
ET NONLITER
SQ SEQUENCE 246 AA; 26884 MW; C9776392C5129A2F CRC64;

Query Match 82.2%; Score 88; DB 2; Length 246;
Best Local Similarity 88.9%; Pred. No. 1.6e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQKQWDGSAI 18
   |||||
DB 29 NTTGVFGLKQKQWDGSAI 46
   |||||

RESULT 5
ID 085817 PRELIMINARY; PRT; 276 AA.
AC 085817;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=SAPULPA;
RA Yu X.-J., Walker D.H.;
RT "Ehrlichia chaffeensis 28 kDa outer membrane protein.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF077734; AAC31547.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30027 MW; 2FD3698FCF1F60BE CRC64;

Query Match 82.2%; Score 88; DB 2; Length 276;
Best Local Similarity 88.9%; Pred. No. 1.9e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQKQWDGSAI 18
   |||||
DB 59 NTTGVFGLKQKQWDGSAI 76
   |||||

RESULT 6
```


O85816
 ID O85816 PRELIMINARY; PRT; 280 AA.
 AC O85816;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DE OUTR MEMBRANE PROTEIN P28 PRECURSOR.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=91HEL7;
 RA Yu X.-J., Walker D.H.;
 RT "Ehrlichia chaffeensis 28 kDa outer membrane protein.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF077732; AAC31545.1; -;
 DR InterPro; IPR002566; Surface_Ag_msp4.
 DR Pfam; PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 280 AA; 30277 MW; 91C54AC78507A63F CRC64;

Query Match 82.2%; Score 88; DB 2; Length 280;
 Best Local Similarity 80.0%; Pred. No. 1.9e-06;
 Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAISN 20
 Db 59 STTAGVFLKQDWDGSAISH 78

RESULT 7
 Q3ZGM9
 ID Q3ZGM9 PRELIMINARY; PRT; 280 AA.
 AC Q3ZGM9;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE OUTR MEMBRANE PROTEIN P28 PRECURSOR.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAX;
 RA Yu X.-J., Walker D.H.;
 RT "Ehrlichia chaffeensis 28 kDa outer membrane protein.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF077733; AAC31546.1; -;
 DR InterPro; IPR002566; Surface_Ag_msp4.
 DR Pfam; PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 280 AA; 30304 MW; 91C54AC7851B77F2 CRC64;

Query Match 82.2%; Score 88; DB 2; Length 280;
 Best Local Similarity 80.0%; Pred. No. 1.9e-06;
 Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAISN 20
 Db 59 STTAGVFLKQDWDGSAISH 78

RESULT 8
 O52107
 ID O52107 PRELIMINARY; PRT; 280 AA.
 AC O52107;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE OMP-1F.

GN OMP-1F.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ARKANSAS;
 RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
 RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis are encoded by a polymorphic multigene family.";
 RL Infect. Immun. 66:132-139(1998).
 DR EMBL; U72291; AAC02940.1; -;
 DR InterPro; IPR002566; Surface_Ag_msp4.
 DR Pfam; PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 280 AA; 30731 MW; CCAA6C34E2AF393E CRC64;

Query Match 82.2%; Score 88; DB 2; Length 280;
 Best Local Similarity 84.2%; Pred. No. 1.9e-06;
 Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAIS 19
 Db 60 NTTGVFGLKQDWDGSTIS 78

RESULT 9
 Q3ZGJ2
 ID Q3ZGJ2 PRELIMINARY; PRT; 288 AA.
 AC Q3ZGJ2;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE 30-KDA MAJOR OUTER MEMBRANE PROTEIN (P28-8).
 GN P30 OR P28-8.
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OKLAHOMA;
 RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
 RT "Cloning and characterization of multigenes encoding the immunodominant 30-kilodalton major outer membrane proteins of Ehrlichia canis and application of the recombinant protein for serodiagnosis.";
 RL J. Clin. Microbiol. 36:2671-2680(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAKE;
 RA McBride J.W., Yu X.J., Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic antigen.";
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAKE;
 RX MEDLINE=20432107; PubMed=10574556;
 RA McBride J.W., Yu X.J., Walker D.H.;
 RT "A conserved, transcriptionally active p28 multigene locus of Ehrlichia canis.";
 RL Gene 254:245-252(2000).
 DR EMBL; AF078553; AAC68667.1; -;
 DR EMBL; AF082744; AAG14362.1; -;
 DR InterPro; IPR002566; Surface_Ag_msp4.
 DR Pfam; PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 288 AA; 31590 MW; 86CAECB88E9BF5E CRC64;

Query Match 76.6%; Score 82; DB 2; Length 288;
 Best Local Similarity 70.0%; Pred. No. 1.9e-05;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNDGSAISN 20
 DB 60 NTTGVGFLKQNDGATIKD 79

RESULT 10

ID Q9F473 PRELIMINARY; PRT; 280 AA.
 AC Q9F473; 2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE P28-6.
 GN P28-6.
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAKE;
 EX MEDLINE=95242757; PubMed=10225842;
 FA McBride J.W., Yu X.J., Walker D.H.;
 FT "Molecular cloning of the gene for a conserved major immunoreactive
 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 antigen";
 RT Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN-JAKE;
 RC MEDLINE=20432107; PubMed=10974556;
 RX McBride J.W., Yu X.J., Walker D.H.;
 RA "A conserved, transcriptionally active p28 multigene locus of
 Ehrlichia canis";
 RT Gene 254:245-252(2000).
 RL ENBL; AF082744; AAG14361.1; -;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1
 SQ SEQUENCE 280 AA; 30762 MW; BE284A4B94FE3123 CRC64;

Query Match 74.8%; Score 80; DB 2; Length 280;
 Best Local Similarity 70.0%; Pred. No. 3.8e-05;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNDGSAISN 20
 DB 59 NSTGVGFLKHDWNGGTISN 78

RESULT 11

ID Q9ADV3 PRELIMINARY; PRT; 280 AA.
 AC Q9ADV3;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE MAJOR OUTER MEMBRANE PROTEIN P30-2.
 GN P30-2.
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OKLAHOMA;
 RX MEDLINE=98371112; PubMed=9705412;

RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
 RT "Cloning and characterization of multigenes encoding the
 immunodominant 30-kilodalton major outer membrane proteins of
 Ehrlichia canis and application of the recombinant protein for
 serodiagnosis";
 RT J. Clin. Microbiol. 36:2671-2680(1998).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN-OKLAHOMA;
 RC MEDLINE=21153566; PubMed=11254561;
 RX Ohashi N., Rikihisa Y., Unver A.;
 RA "Analysis of Transcriptionally Active Gene Clusters of Major Outer
 Membrane Protein Multigene Family in Ehrlichia canis and E.
 chaffeensis";
 RT Infect. Immun. 69:2083-2091(2001).
 RL ENBL; AF078553; AAK28699.1; -;
 DR EMBL; AF078553; AAK28699.1; -;
 SQ SEQUENCE 280 AA; 30803 MW; 27238BE1C7E68A91 CRC64;

Query Match 74.8%; Score 80; DB 2; Length 280;
 Best Local Similarity 70.0%; Pred. No. 3.8e-05;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNDGSAISN 20
 DB 59 NSTGVGFLKHDWNGGTISN 78

RESULT 12

ID O52105 PRELIMINARY; PRT; 286 AA.
 AC O52105;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE OMP-1D.
 GN OMP-1D.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ARKANSAS;
 RX MEDLINE=98084465; PubMed=9423849;
 RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
 RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
 are encoded by a polymorphic multigene family";
 RL Infect. Immun. 66:132-139(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ARKANSAS;
 RX MEDLINE=98321180; PubMed=9647746;
 RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
 RA Alleman A.R.;
 RT "Molecular characterization of a 28 kDa surface antigen gene family of
 the tribe Ehrlichiae"; Commun. 247:636-643(1998).
 RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
 DR ENBL; U72291; AAC02938.1; -;
 DR EMBL; AF062761; AAC26718.1; -;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 286 AA; 31509 MW; FL45A79270F386BE CRC64;

Query Match 68.2%; Score 73; DB 2; Length 286;
 Best Local Similarity 68.4%; Pred. No. 0.00055;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNDGSAIS 19
 DB 60 NTTGVGFLKHDWDRCVIS 78

```
RESULT 13
Q9R8A9 PRELIMINARY; PRT; 278 AA.
AC Q9R8A9;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE 28 KDA OUTER MEMBRANE PROTEIN (FRAGMENT).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LOUISIANA;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of a conserved major immunoreactive 28-kilodalton
RT protein gene from a polymorphic multiple gene family of Ehrlichia
RT canis.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF082745; AAC64551.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 67.3%; Score 72; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.00077;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTVGVFGLKQNWGSAI 18
Db :|||||:|||||
61 STVGVFGLKHDWGSPI 77

RESULT 14
Q9R8A8 PRELIMINARY; PRT; 278 AA.
AC Q9R8A8;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE 28 KDA OUTER MEMBRANE PROTEIN (FRAGMENT).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of a conserved major immunoreactive 28-kilodalton
RT protein gene from a polymorphic multiple gene family of Ehrlichia
RT canis.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF082746; AAC64552.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 67.3%; Score 72; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.00077;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTVGVFGLKQNWGSAI 18
Db :|||||:|||||
61 STVGVFGLKHDWGSPI 77

RESULT 15
Q9R8A7 PRELIMINARY; PRT; 278 AA.
AC Q9R8A7;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE 28 KDA OUTER MEMBRANE PROTEIN (FRAGMENT).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DEMOM (NORTH CAROLINA);
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of a conserved major immunoreactive 28-kilodalton
RT protein gene from a polymorphic multiple gene family of Ehrlichia
RT canis.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF082747; AAC64553.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 67.3%; Score 72; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.00077;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTVGVFGLKQNWGSAI 18
Db :|||||:|||||
61 STVGVFGLKHDWGSPI 77

Search completed: March 14, 2002, 09:24:09
Job time: 980 sec
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:07:40 ; Search time 54.54 Seconds
(without alignments)
7.427 Million cell updates/sec

Title: US-09-765-739A-4
Perfect score: 98
Sequence: 1 NPTVALYGLKQDWNGVSA 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pap.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pap.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pap.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pap.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pap.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	88	89.8	278	4	US-08-953-326-16
2	61	62.2	280	4	US-08-953-326-17
3	57	58.2	276	4	US-08-953-326-18
4	57	58.2	280	3	US-08-733-230-4
5	57	58.2	280	4	US-08-953-326-4
6	51	52.0	287	3	US-08-733-230-2
7	51	52.0	287	4	US-08-953-326-2
8	50	51.0	286	4	US-08-953-326-15
9	46	46.9	133	4	US-08-953-326-20
10	42	42.9	915	4	US-09-346-237-2
11	42	42.9	928	1	US-08-474-140-11
12	42	42.9	928	1	US-08-477-630-11
13	42	42.9	928	1	US-08-472-293-11
14	42	42.9	928	1	US-08-474-545-11
15	42	42.9	928	2	US-08-478-341-11
16	42	42.9	928	3	US-08-996-733-11
17	40	40.8	595	2	US-08-677-049-11
18	39	39.8	263	1	US-07-901-707-7
19	39	39.8	263	1	US-07-988-430-7
20	39	39.8	263	1	US-08-425-336-7
21	39	39.8	263	1	US-08-488-113B-7
22	39	39.8	263	1	US-08-477-484B-7
23	39	39.8	263	2	US-08-646-360-7
24	39	39.8	263	4	US-08-839-765-7
25	39	39.8	263	4	US-09-136-389-7
26	39	39.8	263	5	PCT-US92-09487-7
27	39	39.8	609	1	US-07-798-776-2

28	39	39.8	609	3	US-08-251-288A-2	Sequence 2, Appli
29	39	39.8	609	3	US-09-298-819A-2	Sequence 2, Appli
30	39	39.8	671	2	US-08-737-716-13	Sequence 13, Appl
31	38.5	39.3	368	5	PCT-US93-11703-24	Sequence 24, Appl
32	38.5	39.3	393	1	US-08-499-568-15	Sequence 15, Appl
33	38.5	39.3	393	2	US-08-793-958-15	Sequence 15, Appl
34	38.5	39.3	393	2	US-08-956-998-2	Sequence 2, Appli
35	38.5	39.3	642	1	US-08-706-936-2	Sequence 2, Appli
36	38.5	39.3	643	2	US-08-616-844-39	Sequence 39, Appl
37	38.5	39.3	643	2	US-08-599-654-39	Sequence 39, Appl
38	38.5	39.3	643	3	US-08-944-868A-39	Sequence 39, Appl
39	38.5	39.3	643	3	US-08-944-423A-39	Sequence 39, Appl
40	38.5	39.3	643	3	US-08-944-496-39	Sequence 39, Appl
41	38	38.8	55	1	US-07-869-912-4	Sequence 1, Appli
42	38	38.8	55	2	US-08-369-829A-1	Sequence 1, Appli
43	38	38.8	241	4	US-08-896-933-32	Sequence 32, Appl
44	38	38.8	738	4	US-08-989-385-1	Sequence 1, Appli
45	38	38.8	862	4	US-09-346-237-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-953-326-16
; Sequence 16, Application US/08953326
; Patent No. 6251872

GENERAL INFORMATION:

APPLICANT: Barbet, Anthony F.
APPLICANT: Ganta, Roman R.
APPLICANT: McGuire, Travis C.
APPLICANT: Burridge, Michael J.
APPLICANT: Nyika, Aceme
APPLICANT: Rurangirwa, Fred R.
APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-16

Query Match 89.8%; Score 88; DB 4; Length 278;
Best Local Similarity 83.3%; Pred. No. 1.5e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVSA 18
Db 60 NPTVALYGLKQDWNGVSA 77

RESULT 2
US-08-953-326-17
; Sequence 17, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
APPLICANT: Barbet, Anthony F.
APPLICANT: Ganta, Roman R.
APPLICANT: McGuire, Travis C.
APPLICANT: Burridge, Michael J.
APPLICANT: Nyika, Aceme
APPLICANT: Rurangirwa, Fred R.

```
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-17

Query Match          58.2%; Score 61; DB 4; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.0069;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NPTVALYGLKQDWNG 15
Db 60 NTTGVFGLKQNDWG 74

RESULT 3
US-08-953-326-18
; Sequence 18, Application US/089533326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-18

Query Match          58.2%; Score 57; DB 4; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.033;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NPTVALYGLKQDWNG 15
Db 59 NTTGVFGLKQNDWG 73

RESULT 4
US-08-733-230-4
; Sequence 4, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
```

```
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 6
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-4

Query Match          58.2%; Score 57; DB 3; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.034;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NPTVALYGLKQDWNG 15
Db 60 NTTGVFGLKQNDWG 74

RESULT 5
US-08-953-326-4
; Sequence 4, Application US/089533326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
```

```

; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4

Query Match      58.2%; Score 57; DB 4; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.034;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNGV 15
Db 60 NTTGVFGLKQDNGV 74

RESULT 6
US-08-733-230-2
; Sequence 2, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-2

```

```

Query Match      52.0%; Score 51; DB 3; Length 287;
Best Local Similarity 57.1%; Pred. No. 0.37;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 3 TVALYGLKQDNGV 16
Db 63 TQTVFGLKKDWDGV 76

```

RESULT 7

```

US-08-953-326-2
; Sequence 2, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
US-08-953-326-2

Query Match      52.0%; Score 51; DB 4; Length 287;
Best Local Similarity 57.1%; Pred. No. 0.37;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDNGV 16
Db 63 TQTVFGLKKDWDGV 76

RESULT 8
US-08-953-326-15
; Sequence 15, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-15

Query Match      51.0%; Score 50; DB 4; Length 286;
Best Local Similarity 50.0%; Pred. No. 0.56;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWN 14

```


; STREET: 2000 K Street, N.W., Suite 200
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477.630
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilhem F. Gadiano, Esq.
; REGISTRATION NUMBER: 37,136
; REFERENCE/DOCKET NUMBER: 4121-42
; TELEPHONE: (202) 429-0625
; TELEFAX: (202) 293-1850
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 928 amino acids
; TYPE: amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-477-630-11

Query Match 42.9%; Score 42; DB 1; Length 928;
Best Local Similarity 53.3%; Pred. No. 54;
Matches 8; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 1 NPTVALYGLKQDWNG 15
||:||||:|
Db 641 NPGIALYG--EPWTG 653

RESULT 13
US-08-472-293-11
; Sequence 11, Application US/08472293
; Patent No. 5731174
; GENERAL INFORMATION:
; APPLICANT: DEWEER, PHILIPPE
; APPLICANT: AMORY, ANTOINE
; TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH
; TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS
; TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WILLIAN BRINKS HOFER GILSON & LIONE, P.C.
; STREET: 2000 K Street, N.W., Suite 200
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472.293
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilhem F. Gadiano, Esq.
; REGISTRATION NUMBER: 37,136
; REFERENCE/DOCKET NUMBER: 4121-44
; TELEPHONE: (202) 429-0625

; TELEFAX: (202) 293-1850
; TELEX: 650 383-5605
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 928 amino acids
; TYPE: amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-472-293-11

Query Match 42.9%; Score 42; DB 1; Length 928;
Best Local Similarity 53.3%; Pred. No. 54;
Matches 8; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 1 NPTVALYGLKQDWNG 15
||:||||:|
Db 641 NPGIALYG--EPWTG 653

RESULT 14
US-08-474-545-11
; Sequence 11, Application US/08474545
; Patent No. 5736375
; GENERAL INFORMATION:
; APPLICANT: DEWEER, PHILIPPE
; APPLICANT: AMORY, ANTOINE
; TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH
; TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS
; TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WILLIAN BRINKS HOFER GILSON & LIONE, P.C.
; STREET: 2000 K Street, N.W., Suite 200
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474.545
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilhem F. Gadiano, Esq.
; REGISTRATION NUMBER: 37,136
; REFERENCE/DOCKET NUMBER: 4121-43
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 429-0625
; TELEFAX: (202) 293-1850
; TELEX: 650 383-5605
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 928 amino acids
; TYPE: amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-474-545-11

Query Match 42.9%; Score 42; DB 1; Length 928;
Best Local Similarity 53.3%; Pred. No. 54;
Matches 8; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 1 NPTVALYGLKQDWNG 15
||:||||:|
Db 641 NPGIALYG--EPWTG 653

```

RESULT 15
US-08-478-341-11
; Sequence 11, Application US/08478341
; Patent No. 5817498
; GENERAL INFORMATION:
; APPLICANT: DEWEER, PHILIPPE
; TITLE OF INVENTION: PULLULANASE, MICROORGANISMS WHICH
; TITLE OF INVENTION: PRODUCE IT, PROCESSES FOR THE PREPARATION OF THIS
; TITLE OF INVENTION: PULLULANASE AND THE USES THEREOF
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WILLIAM BRINKS HOFER GILSON & LIONE, P.C.
; STREET: 2000 K Street, N.W., Suite 200
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,341
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilhem F. Gadiano, Esq.
; REGISTRATION NUMBER: 37,136
; REFERENCE/DOCKET NUMBER: 4121-45
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 429-0625
; TELEFAX: (202) 293-1850
; TELEX: 650 383-5605
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 928 amino acids
; TYPE: amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-478-341-11

```

```

Query Match 42.9%; Score 42; DB 2; Length 928;
Best Local Similarity 53.3%; Pred. No. 54;
Matches 8; Conservative 2; Mismatches 3; Indels 2; Gaps 1;
QY 1 NPTVALYGLKQDNG 15
Db 641 NPGIALYG--EPWTG 653

```

Search completed: March 14, 2002, 09:07:40
Job time: 841 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:09:37 ; Search time 111.55 Seconds
(without alignments)
11.953 Million cell updates/sec

Title: US-09-765-739A-4
Perfect score: 98
Sequence: 1 NPTVALYGLKQDWNGVSA 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_ll01.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	100.0	280	20	AAV06945
2	88	89.8	278	19	AAW51093
3	88	89.8	278	20	AAV06947
4	88	89.8	278	21	AAB36187
5	88	89.8	278	22	AAU04197
6	62	63.3	280	20	AAV06962
7	62	63.3	280	21	AAV71479
8	61	62.2	280	19	AAW51094
9	61	62.2	280	21	AAB36188
10	61	62.2	280	22	AAU04198
11	59	60.2	288	20	AAV06959

12	58	59.2	280	20	AAV06948
13	57	58.2	256	20	AAV06942
14	57	58.2	276	19	AAW51095
15	57	58.2	276	21	AAB36189
16	57	58.2	276	22	AAU04199
17	57	58.2	280	19	AAW51089
18	57	58.2	280	21	AAB36183
19	57	58.2	280	22	AAU04193
20	57	58.2	281	20	AAV06943
21	52	53.1	278	21	AAV71477
22	52	53.1	307	20	AAV06961
23	51	52.0	287	19	AAW51088
24	51	52.0	287	21	AAB36182
25	51	52.0	287	22	AAU04192
26	50	51.0	286	19	AAW51092
27	50	51.0	286	20	AAV06946
28	50	51.0	286	21	AAB36186
29	50	51.0	286	22	AAU04196
30	49	50.0	276	20	AAV06964
31	46	46.9	132	22	AAU04201
32	46	46.9	133	19	AAW51097
33	46	46.9	133	21	AAB36191
34	46	46.9	133	21	AAV71480
35	46	46.9	283	21	AAV71478
36	45	45.9	904	20	AAV21976
37	43	43.9	140	21	AAB44680
38	42	42.9	50	21	AAB33357
39	42	42.9	53	21	AAB33358
40	42	42.9	187	21	AAV81775
41	42	42.9	237	21	AAB33257
42	42	42.9	311	22	AAW24509
43	42	42.9	815	22	AAE05695
44	42	42.9	817	22	AAE05694
45	42	42.9	826	20	AAV31748

ALIGNMENTS

RESULT 1
AAV06945
ID AAV06945 standard; Protein; 280 AA.
AC AAV06945;
DT 05-JUL-1999 (first entry)
XX E. chafeensis OMP-1C protein.
DE Outer membrane protein; OMP; Ehrlichia chafeensis; E. canis; P30;
KW detection; dog.
XX Ehrlichia chafeensis.
XX WO9913720-A1.
XX 25-MAR-1999.
XX 18-SEP-1998; 98WO-US19600.
XX 19-SEP-1997; 97US-0059353.
XX (OHIS) UNIV OHIO STATE.
XX Ohashi N, Rikihisa Y;
XX WPI; 1999-254290/21.
XX N-PSDB; AAX34745.
XX Novel outer membrane proteins from Ehrlichia chafeensis and
XX Ehrlichia canis
XX Claim 13; Fig 5B; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX
 SQ Sequence 280 AA;

Query Match 100.0%; Score 98; DB 20; Length 280;
 Best Local Similarity 100.0%; Pred. No. 6,7e-09;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVSA 18
 |||||
 Db 60 nptvalyglkqdwngvsa 77

RESULT 2
 AAW51093
 ID AAW51093 standard; Protein; 278 AA.
 AC AAW51093;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Ehrlichia chaffeensis VSA3 protein.
 XX
 KW MAP1 homologue; variable surface antigen; VSA3; rickettsia;
 KW DNA vaccine.
 XX
 OS Ehrlichia chaffeensis.
 XX
 FH Key
 FT Peptide 1..25
 FT /note= "putative signal peptide"
 XX
 PN W09816554-A1.
 XX
 PD 23-APR-1998.
 XX
 PF 17-OCT-1997; 97WO-US19044.
 XX
 PR 17-OCT-1996; 96US-073230.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa PR;
 XX
 DR WPI; 1998-251232/22.
 DR N-PSDB; AAV07179.
 XX
 PT Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 PS
 PS Claim 3; Fig 2A-B; 39pp; English.
 CC This is the full-length variable surface antigen VSA3 protein of
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
 CC partial open reading frame (ORF5) of a genomic locus (see AAV07179)
 CC of E. chaffeensis that was obtained on the basis of homology to the
 CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
 CC This genomic locus included 5 ORFs encoding similar, but
 CC non-identical proteins (see AAW51091-95). A claimed composition
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
 CC (see AAW51088-99) that elicits a protective immune response against a
 CC rickettsial pathogen. The nucleic acid is used, in human or
 CC veterinary medicine, in vaccines to protect against Rickettsia,
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic

CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).
 XX
 SQ Sequence 278 AA;

Query Match 89.8%; Score 88; DB 19; Length 278;
 Best Local Similarity 83.3%; Pred. No. 3.6e-07;
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVSA 18
 |||||
 Db 60 nptvalyglkqdwngvsa 77

RESULT 3
 AAY06947
 ID AAY06947 standard; Protein; 278 AA.
 XX
 AC AAY06947;
 XX
 DT 05-JUL-1999 (first entry)
 XX
 DE E. chaffeensis OMP-1E protein.
 XX
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN W09913720-A1.
 XX
 PD 25-MAR-1999.
 XX
 PF 18-SEP-1998; 98WO-US19600.
 XX
 PR 19-SEP-1997; 97US-0059353.
 XX
 PA (OHIS) UNIV OHIO STATE.
 XX
 PI Ohashi N, Rikihisa Y;
 XX
 DR WPI; 1999-254290/21.
 DR N-PSDB; AAX34747.
 XX
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis
 XX
 PS Claim 15; Fig 7B; 55pp; English.
 CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX
 SQ Sequence 278 AA;

Query Match 89.8%; Score 88; DB 20; Length 278;
 Best Local Similarity 83.3%; Pred. No. 3.6e-07;
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVSA 18
 |||||
 Db 60 nptvalyglkqdwngvsa 77

RESULT 4
 AAB36187
 ID AAB36187 standard; Protein; 278 AA.
 XX

```

AC AAB36187;
XX
DT 02-MAR-2001 (first entry)
XX
DE Ehrlichia chaffeensis partial VSA3.
XX
KW Ehrlichia chaffeensis; VSA3; variable surface antigen 3; MAP1;
KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
KW Rickettsia; Cowdria; Anaplasma; map2; 1hworf3; 4hworf1; 18hworf1;
KW 3gdorf3.
XX
OS Ehrlichia chaffeensis.
XX
PN WO200065063-A2.
XX
PD 02-NOV-2000.
XX
PF 21-APR-2000; 2000WO-US10886.
XX
PR 22-APR-1999; 99US-0130725.
XX
PA (UYFL ) UNIV FLORIDA.
XX
PI Barbet AF, Bowie MV, Ganta RR, Burrridge MJ, Mahan SM, McGuire TC;
PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
XX
DR WPI; 2000-679675/66.
DR N-PSDB; AAC68704.
XX
XX New polynucleotides useful as DNA vaccines for conferring immunity to
PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
PT .
XX
PS Claim 3; Page 44-45; 63pp; English.
XX
CC The present sequence shows a high degree of similarity to the major
CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
CC used in a vaccines to protect animals or humans against rickettsial
CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
CC protective against the rickettsial pathogen. The nucleic acid vaccines
CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
CC Cowdria ruminantium genes designated map 2, 1hworf3, 4hworf1, 18hworf1
CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
CC polypeptides are useful for detecting antibodies associated with
CC infection by a rickettsial pathogen whilst the polynucleotides may be
CC used to detect the presence of rickettsial nucleic acids.
XX
SQ Sequence 278 AA;

Query Match 89.8%; Score 88; DB 21; Length 278;
Best Local Similarity 83.3%; Pred. No. 3.6e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVSA 18
Db |||||||||||1:1:
60 nptvalyglkqdwngvss 77

RESULT 5
AAU04197
ID AAU04197 standard; Protein; 278 AA.
XX
AC AAU04197;
XX
DT 23-OCT-2001 (first entry)
XX
DE Variable surface antigen 3 (VSA3) from Ehrlichia chaffeensis.
XX
KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
KW infection; heartwater; diagnostic; variable surface antigen; VSA.

```

```

XX
OS Ehrlichia chaffeensis.
XX
PN US6251872-B1.
XX
PD 26-JUN-2001.
XX
PF 17-OCT-1997; 97US-0953326.
XX
PR 17-OCT-1996; 96US-0733230.
XX
PA (UYFL ) UNIV FLORIDA.
XX
PI Barbet AF, Ganta RR, McGuire TC, Burrridge MJ, Nyika A;
PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX
DR WPI; 2001-424487/45.
DR N-PSDB; AA507578.
XX
XX New MAP2 genes and polypeptides useful as vaccines for conferring
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
PT as molecular markers in nucleic acid analysis procedures
XX
PS Example 3; Fig 2A-2B; 30pp; English.
XX
CC The sequence represents the amino acid sequence of variable surface
CC antigen 3 (VSA3) isolated from Ehrlichia chaffeensis, which
CC has similarity to major antigen protein (MAP). The MAP polynucleotides
CC and polypeptides are useful as vaccines for conferring immunity to
CC rickettsia infection, including Cowdria ruminantium causing heartwater.
CC The MAP polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures, and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish infectious
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
XX
SQ Sequence 278 AA;

Query Match 89.8%; Score 88; DB 22; Length 278;
Best Local Similarity 83.3%; Pred. No. 3.6e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVSA 18
Db |||||||||||1:1:
60 nptvalyglkqdwngvss 77

RESULT 6
AAU06962
ID AAU06962 standard; Protein; 280 AA.
XX
AC AAU06962;
XX
DT 05-JUL-1999 (first entry)
XX
DE E. canis p30-2 protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; p30;
KW detection; dog.
XX
OS Ehrlichia canis.
XX
PN WO9913720-A1.
XX
PD 25-MAR-1999.
XX
PF 18-SEP-1998; 98WO-US19600.
XX

```

PR 19-SEP-1997; 97US-0059353.

XX (OHIS) UNIV OHIO STATE.

XX Ohashi N, Rikihisa Y;

XX WPI: 1999-254290/21.

DR N-PSDB; AAX34762.

XX Novel outer membrane proteins from Ehrlichia chaffeensis and

XX Ehrlichia canis

XX Disclosure; Fig 22B; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from

XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part

XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown

XX in AY06943-958, the E. canis proteins form part of the p30 family and

XX consist of proteins shown in AAY06959-970. The proteins and genes are

XX used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 280 AA;

SQ

Query Match 63.3%; Score 62; DB 20; Length 280;

Best Local Similarity 66.7%; Pred. No. 0.011;

Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15

DB 59 nstvgvfglkhdwng 73

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

PT ehrlichiosis in dogs -

XX Claim 12; Page 68-69; 86pp; English.

XX The patent relates to homologous 28-kilobalton (kDa) protein genes of

XX Ehrlichia canis, designated Eca28SA1, Eca28SA2, Eca28SA3, Eca28-1 and

XX Eca28-2. These genes are members of a polymorphic multiple gene family

XX and contained in a single locus of 5.592 kb. The 28-kDa proteins are

XX immunoreactive with anti-E. canis serum hence are important

XX immunoprotective antigens. The protein is useful for vaccinating

XX against E. canis infections such as canine ehrlichiosis in dogs.

XX Canine ehrlichiosis, also known as canine tropical pancytopenia, is a

XX tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst

XX different strains of E. canis and hence useful for serodiagnosis of

XX canine ehrlichiosis. The present sequence is a E. canis

XX Eca28SA3 30-kDa protein which is post-translationally modified to a

XX mature 28-kDa protein by cleavage of N-terminal signal sequence.

XX Sequence 280 AA;

SQ

Query Match 63.3%; Score 62; DB 21; Length 280;

Best Local Similarity 66.7%; Pred. No. 0.011;

Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15

DB 59 nstvgvfglkhdwng 73

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

PR 19-SEP-1997; 97US-0059353.

XX (OHIS) UNIV OHIO STATE.

XX Ohashi N, Rikihisa Y;

XX WPI: 1999-254290/21.

DR N-PSDB; AAX34762.

XX Novel outer membrane proteins from Ehrlichia chaffeensis and

XX Ehrlichia canis

XX Disclosure; Fig 22B; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from

XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part

XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown

XX in AY06943-958, the E. canis proteins form part of the p30 family and

XX consist of proteins shown in AAY06959-970. The proteins and genes are

XX used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 280 AA;

SQ

Query Match 63.3%; Score 62; DB 20; Length 280;

Best Local Similarity 66.7%; Pred. No. 0.011;

Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15

DB 59 nstvgvfglkhdwng 73

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

PT ehrlichiosis in dogs -

XX Claim 12; Page 68-69; 86pp; English.

XX The patent relates to homologous 28-kilobalton (kDa) protein genes of

XX Ehrlichia canis, designated Eca28SA1, Eca28SA2, Eca28SA3, Eca28-1 and

XX Eca28-2. These genes are members of a polymorphic multiple gene family

XX and contained in a single locus of 5.592 kb. The 28-kDa proteins are

XX immunoreactive with anti-E. canis serum hence are important

XX immunoprotective antigens. The protein is useful for vaccinating

XX against E. canis infections such as canine ehrlichiosis in dogs.

XX Canine ehrlichiosis, also known as canine tropical pancytopenia, is a

XX tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst

XX different strains of E. canis and hence useful for serodiagnosis of

XX canine ehrlichiosis. The present sequence is a E. canis

XX Eca28SA3 30-kDa protein which is post-translationally modified to a

XX mature 28-kDa protein by cleavage of N-terminal signal sequence.

XX Sequence 280 AA;

SQ

Query Match 63.3%; Score 62; DB 21; Length 280;

Best Local Similarity 66.7%; Pred. No. 0.011;

Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15

DB 59 nstvgvfglkhdwng 73

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

partial open reading frame (ORF4) of a genomic locus (see AAV07179) of *E. chaffeensis* that was obtained on the basis of homology to the major antigenic protein MAP1 (see AAW51088) of *Cowdria ruminantium*. This genomic locus included 5 ORFs encoding similar, but non-identical proteins (see AAW51091-95). A claimed composition comprises a nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99) that elicits a protective immune response against a rickettsial pathogen. The nucleic acid is used, in human or veterinary medicine, in vaccines to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic polypeptides can be used diagnostically to detect antibodies associated with Ehrlichia infection (claimed).

SQ Sequence 280 AA;
Query Match 62.2%; Score 61; DB 19; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.017;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NPTVALYGLKQDWNG 15
| | : : ||||| : |
Db 60 nttigvfglkqdwg 74

RESULT 9
AAB36188
ID AAB36188 standard; Protein; 280 AA.

AC AAB36188;

DT 02-MAR-2001 (first entry)
DE Ehrlichia chaffeensis partial VSA4.

XX Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAP1;
KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
KW 3gdorf3.

XX Ehrlichia chaffeensis.

XX WO200065063-A2.

PN 02-NOV-2000.

PD 21-APR-2000; 2000WO-US10886.

PF 22-APR-1999; 99US-0130725.

PR (UYFL) UNIV FLORIDA.

XX Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;

XX WPI; 2000-679675/66.

DR N-PSDB; AAC68705.

XX New polynucleotides useful as DNA vaccines for conferring immunity to
PT rickettsial infection e.g. heartwater caused by *Cowdria ruminantium*,
PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

PS Claim 3; Page 45-46; 63pp; English.

XX The present sequence shows a high degree of similarity to the major
CC antigenic protein 1 (MAP1) of *Ehrlichia* sp. The MAP1 gene may be
CC used in a vaccines to protect animals or humans against rickettsial
CC diseases caused by a organisms of *Rickettsia* sp., *Ehrlichia* sp.,
CC *Anaplasma* sp. or *Cowdria* sp. The vaccine elicits an immune response
CC protective against the rickettsial pathogen. The nucleic acid vaccines
CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
CC *Cowdria ruminantium* genes designated map 2, lhworf3, 4hworf1, 18hworf1

CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
CC polypeptides are useful for detecting antibodies associated with
CC infection by a rickettsial pathogen whilst the polynucleotides may be
CC used to detect the presence of rickettsial nucleic acids.

XX SQ Sequence 280 AA;

Query Match 62.2%; Score 61; DB 21; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.017;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NPTVALYGLKQDWNG 15
| | : : ||||| : |
Db 60 nttigvfglkqdwg 74

RESULT 10
AAU04198
ID AAU04198 standard; Protein; 280 AA.

AC AAU04198;

DT 23-OCT-2001 (first entry)

DE Variable surface antigen 4 (VSA4) from *Ehrlichia chaffeensis*.

XX Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
KW infection; heartwater; diagnostic; variable surface antigen; VSA.

XX Ehrlichia chaffeensis.

XX US6251872-B1.

PN 26-JUN-2001.

PD 17-OCT-1997; 97US-0953326.

PF 17-OCT-1996; 96US-0733230.

PR (UYFL) UNIV FLORIDA.

XX Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;

XX WPI; 2001-424487/45.

DR N-PSDB; AAS07578.

XX New MAP2 genes and polypeptides useful as vaccines for conferring
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
PT as molecular markers in nucleic acid analysis procedures -

PS Example 3; Fig 2A-2B; 30pp; English.

XX The sequence represents the amino acid sequence of variable surface
CC antigen 4 (VSA4) isolated from *Ehrlichia chaffeensis*, which
CC has similarity to major antigen protein (MAP). The MAP polynucleotides
CC and polypeptides are useful as vaccines for conferring immunity to
CC rickettsia infection, including *Cowdria ruminantium* causing heartwater.
CC The MAP polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures, and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish infectious
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
CC needed.

XX SQ Sequence 280 AA;

Query Match 62.2%; Score 61; DB 22; Length 280;

```

Best Local Similarity 60.0%; Pred. No. 0.017;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
   | | : : ||||| : |
DB 60 ntgtgvgfkgqdwgdg 74

RESULT 11
AAY06959
ID AAY06959 standard; Protein; 288 AA.
AC AAY06959;
DT 05-JUL-1999 (first entry)
DE E. canis P30 protein.
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
OS Ehrlichia canis.
PN WO9913720-A1.
PD 25-MAR-1999.
PF 18-SEP-1998; 98WO-US19600.
PR 19-SEP-1997; 97US-0059353.
PA (OHIS ) UNIV OHIO STATE.
PI Ohashi N, Rikihisa Y;
DR WPI; 1999-254290/21.
DR N-PSDB; AAX34759.
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
XX Ehrlichia canis
XX Claim 16; Fig 19B; 55pp; English.
XX The invention provides isolated outer membrane proteins (OMP) from
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
XX in AAY06943-958. The E. canis proteins form part of the P30 family and
XX consist of proteins shown in AAY06959-970. The proteins and genes are
XX used to detect E. chaffeensis in patients and E. canis in dogs.
XX Sequence 288 AA;

Query Match 60.2%; Score 59; DB 20; Length 288;
Best Local Similarity 52.9%; Pred. No. 0.039;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVS 17
   | | : : ||||| : |
DB 60 ntgtgvgfkgqdwgdg 76

RESULT 12
AAY06948
ID AAY06948 standard; Protein; 280 AA.
AC AAY06948;
DT 05-JUL-1999 (first entry)
DE E. chaffeensis OMP-1F protein.
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;

```

```

KW detection; dog.
XX Ehrlichia chaffeensis.
OS WO9913720-A1.
PN 25-MAR-1999.
PD 18-SEP-1998; 98WO-US19600.
PF 19-SEP-1997; 97US-0059353.
PR (OHIS ) UNIV OHIO STATE.
PA Ohashi N, Rikihisa Y;
DR WPI; 1999-254290/21.
DR N-PSDB; AAX34748.
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
XX Ehrlichia canis
XX Claim 16; Fig 8B; 55pp; English.
XX The invention provides isolated outer membrane proteins (OMP) from
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
XX in AAY06943-958. The E. canis proteins form part of the P30 family and
XX consist of proteins shown in AAY06959-970. The proteins and genes are
XX used to detect E. chaffeensis in patients and E. canis in dogs.
XX Sequence 280 AA;

Query Match 59.2%; Score 58; DB 20; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.057;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
   | | : : ||||| : |
DB 60 ntgtgvgfkgqdwgdg 74

RESULT 13
AAY06942
ID AAY06942 standard; Protein; 256 AA.
AC AAY06942;
DT 05-JUL-1999 (first entry)
DE E. chaffeensis p28 protein.
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
OS Ehrlichia chaffeensis.
PN WO9913720-A1.
PD 25-MAR-1999.
PF 18-SEP-1998; 98WO-US19600.
PR 19-SEP-1997; 97US-0059353.
PA (OHIS ) UNIV OHIO STATE.
PI Ohashi N, Rikihisa Y;
DR WPI; 1999-254290/21.
DR N-PSDB; AAX34742.
XX

```


PT Novel outer membrane proteins from Ehrlichia chaffeensis and
XX Ehrlichia canis

PS Claim 18; Fig 1; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the p30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX SQ Sequence 256 AA;

Query Match 58.2%; Score 57; DB 20; Length 256;
Best Local Similarity 60.0%; Pred. No. 0.076; Length 256;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NPTVALYGLKQDWNG 15
| | | :| | | :| |
Db 34 nttvgvfglkqndg 48

RESULT 14

AAW51095
ID AAW51095 standard; Protein; 276 AA.

AC AAW51095;

DT 14-SEP-1998 (first entry)

DE Ehrlichia chaffeensis VSA5 protein (partial sequence).

KW MAP1 homologue; variable surface antigen; VSA5; rickettsia;
XX DNA vaccine.

OS Ehrlichia chaffeensis.

EH Key Location/Qualifiers

FT Peptide 1..25 /note= "putative signal peptide"

XX WO9816554-A1.

XX 23-APR-1998.

PF 17-OCT-1997; 97WO-US19044.

PR 17-OCT-1996; 96US-0733230.

XX (UYFL) UNIV FLORIDA.

XX Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;

PI Nyika A, Rurangirwa FR;

XX WPI; 1998-251232/22.

DR N-PSDB; AAV07179.

XX Composition containing nucleic acid encoding rickettsial antigen -
PT useful for, e.g. stimulating protective immune response in humans or
PT animals

XX Claim 3; Fig 2B; 39pp; English.

XX This is the near full-length variable surface antigen VSA5 protein
CC of Ehrlichia chaffeensis; it lacks about 5-7 C-terminal amino acid
CC residues. The VSA5 amino acid sequence was deduced from a partial
CC open reading frame (ORF5) of a genomic locus (see AAV07179) of E.
CC chaffeensis that was obtained on the basis of homology to the major
CC antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium. This
CC genomic locus included 5 ORFs encoding similar, but non-identical
CC proteins (see AAW51091-95). A claimed composition comprises a

CC nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99)
CC that elicits a protective immune response against a rickettsial
CC pathogen. The nucleic acid is used, in human or veterinary
CC medicine, in vaccines to protect against Rickettsia, Ehrlichia,
CC Anaplasma and Cowdria species. The Ehrlichia antigenic
CC polypeptides can be used diagnostically to detect antibodies
CC associated with Ehrlichia infection (claimed).

XX SQ Sequence 276 AA;

Query Match 58.2%; Score 57; DB 19; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.083;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NPTVALYGLKQDWNG 15
| | | :| | | :| |
Db 59 nttvgvfglkqndg 73

RESULT 15

AAB36189
ID AAB36189 standard; Protein; 276 AA.

AC AAB36189;

DT 02-MAR-2001 (first entry)

XX Ehrlichia chaffeensis partial VSA5.

XX Ehrlichia chaffeensis; VSA5; variable surface antigen 5; MAP1;
KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
KW 3gdorf3.

OS Ehrlichia chaffeensis.

XX WO2000065063-A2.

XX 02-NOV-2000.

PF 21-APR-2000; 2000WO-US10886.

PR 22-APR-1999; 99US-0130725.

XX (UYFL) UNIV FLORIDA.

XX Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;

XX WPI; 2000-679675/66.

DR N-PSDB; AAC68706.

XX New polynucleotides useful as DNA vaccines for conferring immunity to
PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

XX Claim 3; Page 47; 63pp; English.

XX The present sequence shows a high degree of similarity to the major
CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
CC used in a vaccines to protect animals or humans against rickettsial
CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
CC protective against the rickettsial pathogen. The nucleic acid vaccines
CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
CC polypeptides are useful for detecting antibodies associated with
CC infection by a rickettsial pathogen whilst the polynucleotides may be
XX used to detect the presence of rickettsial nucleic acids.

SQ Sequence 276 AA;

Query Match 58.2%; Score 57; DB 21; Length 276;
 Best Local Similarity 60.0%; Pred. No. 0.083;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
 | | | : | | | : | |
 Db 59 ntivgvfglkqwdg 73

Search completed: March 14, 2002, 09:09:37
 Job time: 428 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:24:09 ; Search time 103.37 Seconds
(without alignments)
26.886 Million cell updates/sec

Title: US-09-765-739A-5
Perfect score: 105
Sequence: 1 NTTVGFGIEQWDRCVIS 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phase:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	286	2	052105 ehrlichia c
2	77	73.3	280	2	085357 ehrlichia c
3	74	70.5	280	2	052107 ehrlichia c
4	73	69.5	276	2	085358 ehrlichia c
5	73	69.5	281	2	09AC19 ehrlichia c
6	71	67.6	288	2	09ZGJ2 ehrlichia c
7	70	66.7	246	2	09RH35 ehrlichia c
8	70	66.7	276	2	09RH35 ehrlichia c
9	69	65.7	280	2	085816 ehrlichia c
10	69	65.7	280	2	09ZGJ2 ehrlichia c
11	62	59.0	280	2	09F473 ehrlichia c
12	62	59.0	280	2	09ADV3 ehrlichia c
13	58	55.2	278	2	09R8A9 ehrlichia c
14	58	55.2	278	2	09R8A8 ehrlichia c
15	58	55.2	278	2	09R8A7 ehrlichia c
16	58	55.2	278	2	09R8A6 ehrlichia c
17	58	55.2	278	2	09R8A5 ehrlichia c
18	58	55.2	278	2	09R3J3 ehrlichia c
19	58	55.2	278	2	09F472 ehrlichia c

20	58	55.2	307	2	09ZGJ1	Q9ZGJ1 ehrlichia c
21	54	51.4	276	2	09F475	Q9F475 ehrlichia c
22	51	48.6	278	2	052106	052106 ehrlichia c
23	50	47.6	280	2	052104	052104 ehrlichia c
24	47	44.8	160	2	09RW10	Q9RW10 deinococcus
25	46	43.8	272	2	09AMF6	Q9AMF6 cowdria sp.
26	46	43.8	284	2	046327	Q46327 cowdria rum
27	46	43.8	284	2	09AFAL	Q9AFAL cowdria rum
28	45	42.9	290	2	09AEU3	Q9AEU3 cowdria rum
29	44.5	42.4	91	2	09L5P8	Q9L5P8 salmonella
30	44	41.9	172	2	09CKK3	Q9CKK3 pasteurella
31	44	41.9	265	2	09AF99	Q9AF99 cowdria rum
32	44	41.9	270	2	09AF98	Q9AF98 cowdria rum
33	44	41.9	287	2	046329	Q46329 cowdria rum
34	44	41.9	287	2	046331	Q46331 cowdria rum
35	44	41.9	287	2	09R425	Q9R425 cowdria rum
36	44	41.9	290	2	046324	Q46324 cowdria rum
37	44	41.9	290	2	046330	Q46330 cowdria rum
38	44	41.9	290	2	046332	Q46332 cowdria rum
39	44	41.9	290	2	046333	Q46333 cowdria rum
40	44	41.9	898	11	008721	008721 rattus norv
41	43	41.0	281	2	046328	Q46328 cowdria rum
42	43	41.0	281	2	Q9S6H1	Q9S6H1 cowdria rum
43	43	41.0	281	2	Q9S6H0	Q9S6H0 cowdria rum
44	43	41.0	367	2	Q9XD58	Q9XD58 pseudomonas
45	42.5	40.5	417	10	Q9SIS4	Q9SIS4 arabidopsis

ALIGNMENTS

RESULT 1
O52105 PRELIMINARY; PRT: 286 AA.
ID O52105
AC O52105;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE OMP-1D.
GN OMP-1D.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikhiisa Y.,
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
are encoded by a polymorphic multigene family";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
the tribe Ehrlichieae";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL; U72291; AAC26718.1;
DR EMBL; AF062761; AAC26718.1;
DR InterPro; IP0002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 286 AA; 31509 MW; F145A79270F386BE CRC64;

Query Match 100.0% Score 105; DB 2; Length 286;
Best Local Similarity 100.0%; Pred. No. 3.le-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 NTTVGFGIEQWDRCVIS 19

Db 60 NTTGVFGIEQDWDRCVIS 78
|||||

RESULT 2
ID O85357 PRELIMINARY; PRT; 280 AA.
AC O85357;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 28 KDA MAJOR SURFACE ANTIGEN-4.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burrigge M.J.,
RA Alleman A.R.;
RT "Molecular characterization of a 28 kda surface antigen gene family of
the tribe Ehrlichiae.";
RL - Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL; AF062761; AAC26720.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 280 AA; 30743 MW; FBB841DAF08EE4DC CRC64;
SQ SEQUENCE 280 AA; 30743 MW; FBB841DAF08EE4DC CRC64;

Query Match 73.3%; Score 77; DB 2; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.00011;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
|||:||||:|||||
Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 3
ID O52107 PRELIMINARY; PRT; 280 AA.
AC O52107;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE OMP-1F.
GN OMP-1F.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
DR EMBL; U72291; AAC02940.1;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30731 MW; CCAA6C34E2AF393E CRC64;

Query Match 70.5%; Score 74; DB 2; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.00035;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
||| |||||:|||||

Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 4
ID O85358 PRELIMINARY; PRT; 276 AA.
AC O85358;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 28 KDA MAJOR SURFACE ANTIGEN-5 (FRAGMENT).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burrigge M.J.,
RA Alleman A.R.;
RT "Molecular characterization of a 28 kda surface antigen gene family of
the tribe Ehrlichiae.";
RL - Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL; AF062761; AAC26716.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 276 AA; 29782 MW; A9AB2A92263CA4EA CRC64;
SQ SEQUENCE 276 AA; 29782 MW; A9AB2A92263CA4EA CRC64;

Query Match 69.5%; Score 73; DB 2; Length 276;
Best Local Similarity 68.4%; Pred. No. 0.0005;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
|||||:|||||
Db 59 NTTGVFGIEQDWDRCVIS 77

RESULT 5
ID O9AC19 PRELIMINARY; PRT; 281 AA.
AC O9AC19;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN P28.
GN P28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
DR EMBL; U72291; AAC02940.1;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
Membrane Protein Multigene Family in Ehrlichia canis and E.
chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; U72291; AAC28673.1; -
SQ SEQUENCE 281 AA; 30343 MW; A99E5F7C4459AA9A CRC64;

Query Match 69.5%; Score 73; DB 2; Length 281;
Best Local Similarity 68.4%; Pred. No. 0.00052;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCV18
DB 59 NTTGVFGLKQNDGSAI 77

RESULT 6
Q9ZGJ2 PRELIMINARY; PRT; 288 AA.
ID Q9ZGJ2;
AC Q9ZGJ2;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 30-KDA MAJOR OUTER MEMBRANE PROTEIN (P28-8).
GN P30 OR P28-8.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis."
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis."
RL Gene 254:245-252(2000).
DR EMBL; AF078553; AAC68667.1; -;
DR EMBL; AF082744; AAG14362.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 288 AA; 31590 MW; 86DCAECB88E9BF5E CRC64;

Query Match 67.6%; Score 71; DB 2; Length 288;
Best Local Similarity 66.7%; Pred. No. 0.0011;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCV18
DB 60 NTTGVFGLKQNDGATI 77

RESULT 7
Q9RH35 PRELIMINARY; PRT; 246 AA.
ID Q9RH35
AC Q9RH35;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE OUTER MEMBRANE PROTEIN P28 PRECURSOR (FRAGMENT).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ST.VINCENT;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis."
RL J. Clin. Microbiol. 37:1137-1143(1999).
DR EMBL; AF077735; AAC31548.1; -;
DR InterPro: IPR001702; Gram_neg_dorin.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
DR PRINTS; PR00182; ECOLNEIPORIN.
FT NON_TER 1
SQ SEQUENCE 246 AA; 26884 MW; C9776392C5129A2F CRC64;

Query Match 66.7%; Score 70; DB 2; Length 246;
Best Local Similarity 66.7%; Pred. No. 0.0014;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCV18
DB 29 NTTAGVFLKQNDGSAI 46

RESULT 8
O85817 PRELIMINARY; PRT; 276 AA.
ID O85817
AC O85817;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAPULPA;
RA Yu X.-J., Walker D.H.;
RT "Ehrlichia chaffeensis 28 kda outer membrane protein."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077734; AAC31547.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30027 MW; 2FD3698FCF1F60BE CRC64;

Query Match 66.7%; Score 70; DB 2; Length 276;
Best Local Similarity 66.7%; Pred. No. 0.0016;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCV18
DB 59 NTTAGVFLKQNDGSAI 76

RESULT 9
O85816 PRELIMINARY; PRT; 280 AA.
ID O85816
AC O85816;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR.
OS Ehrlichia chaffeensis.

RESULT 13

Q9R8A9
ID Q9R8A9 PRELIMINARY; PRT; 278 AA.

AC Q9R8A9;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE 28 KDA OUTER MEMBRANE PROTEIN (FRAGMENT).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-LOUISIANA;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of a conserved major immunoreactive 28-kilodalton protein gene from a polymorphic multiple gene family of Ehrlichia canis."
RT Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF082745; AAC64551.1;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match

Best Local Similarity 55.2%; Score 58; DB 2; Length 278;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTGVGFGIEQDWD 14
:|||||::|||
Db 61 STVGVGLKHDWD 73

RESULT 14

Q9R8A8
ID Q9R8A8 PRELIMINARY; PRT; 278 AA.

AC Q9R8A8;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE 28 KDA OUTER MEMBRANE PROTEIN (FRAGMENT).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-OKLAHOMA;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of a conserved major immunoreactive 28-kilodalton protein gene from a polymorphic multiple gene family of Ehrlichia canis."
RT Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF082746; AAC64552.1;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match

Best Local Similarity 55.2%; Score 58; DB 2; Length 278;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTGVGFGIEQDWD 14
:|||||::|||
Db 61 STVGVGLKHDWD 73

RESULT 15

Q9R8A7
ID Q9R8A7 PRELIMINARY; PRT; 278 AA.

AC Q9R8A7;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE 28 KDA OUTER MEMBRANE PROTEIN (FRAGMENT).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-DEMOM (NORTH CAROLINA);
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of a conserved major immunoreactive 28-kilodalton protein gene from a polymorphic multiple gene family of Ehrlichia canis."
RT Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF082747; AAC64553.1;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match

Best Local Similarity 55.2%; Score 58; DB 2; Length 278;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTGVGFGIEQDWD 14
:|||||::|||
Db 61 STVGVGLKHDWD 73

Search completed: March 14, 2002, 09:24:10

Job time: 981 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:24:51 ; Search time 32.82 Seconds
(without alignments)
21.226 Million cell updates/sec

Title: US-09-765-739A-5

Perfect score: 105

Sequence: 1 NTTGVFGIEQDWDRCVIS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	44.5	42.4	2182	1 POLG_CXB1J	P08291 c genome po
2	42	40.0	171	1 SORC_SCHJA	Q94743 schistosoma
3	41	39.0	572	1 DPV4_MOUSE	O35098 mus musculus
4	41	39.0	573	1 MD13_PRUSE	P52707 prunus sero
5	41	39.0	576	1 MDL2_PRUSE	O50048 prunus sero
6	40.5	38.6	1790	1 VIT_ANTGR	Q05808 anthomus
7	40	38.1	46	1 TXAL_ANESU	P01533 anemonia su
8	40	38.1	142	1 FER_THEAC	P00218 thermoplas
9	40	38.1	463	1 FLGE_TREPH	Q56326 treponema p
10	40	38.1	540	1 HXTD_YEAST	P42833 saccharomyc
11	40	38.1	564	1 DPV4_RAT	Q62951 rattus norv
12	40	38.1	1012	1 MA2B_MOUSE	O09159 mus musculus
13	39	37.1	279	1 NRTB_PHOLA	Q51881 phormidium
14	39	37.1	361	1 MLTB_ECOLI	P41052 escherichia
15	39	37.1	371	1 Y028_ARCFU	O30207 archaeoglob
16	39	37.1	377	1 YA67_METTH	O27139 methanobact
17	39	37.1	397	1 P2X3_HUMAN	P56373 homo sapien
18	39	37.1	397	1 P2X3_RAT	P49654 rattus norv
19	39	37.1	548	1 SYPB_AERPE	O9Y913 aeropyrum p
20	39	37.1	751	1 YLX5_CAEEL	P46501 caenorhabdi
21	39	37.1	993	1 AFSR_STRCO	P25941 streptomyce
22	39	37.1	1033	1 S190_YEAST	P36123 saccharomyc
23	38.5	36.7	223	1 FLGA_AQUAE	O67005 aquifex aeo
24	38.5	36.7	644	1 SYR_AERPE	O9YB39 aeropyrum p
25	38	36.2	282	1 BIOB_HELPJ	Q9ZJX8 helicobacte
26	38	36.2	282	1 BIOB_HELPJ	O25956 helicobacte
27	38	36.2	353	1 DHYS_NEUCR	P49365 neurospora
28	38	36.2	430	1 Y588_METJA	Q58008 methanococ
29	38	36.2	460	1 V51K_ACLSP	P27739 apple chlor
30	38	36.2	608	1 EX5A_ECOLI	P04993 escherichia
31	38	36.2	668	1 UVRB_TREPA	O83154 treponema p
32	38	36.2	729	1 FHUE_ECOLI	P16869 escherichia
33	38	36.2	1023	1 CLOC_DROME	O61735 drosophila

RESULT 1

ID	POLG_CXB1J	STANDARD;	PRT;	2182 AA.
AC	P08291.			
DT	01-AUG-1988 (Rel. 08, Created)			
DT	01-AUG-1988 (Rel. 08, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	GENOME POLYPROTEIN [CONTAINS: COAT PROTEIN VP4 (P1A); COAT PROTEIN VP2 (P1B); COAT PROTEIN VP3 (P1C); COAT PROTEIN VP1 (PID); PICORNAIN 2A (EC 3.4.22.29) (P2A); CORE PROTEIN P2B; CORE PROTEIN P2C; CORE PROTEIN P3A; GENOME-LINKED PROTEIN VP3 (P3B); PICORNAIN 3C (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (P3D)].			
OS	Coxsackievirus B1 (strain Japan).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;			
OC	Enterovirus.			
OX	NCBI_TaxID=103902;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87122157; PubMed=3027969;			
RA	Iizuka N., Kuge S., Nomoto A.;			
RT	"Complete nucleotide sequence of the genome of coxsackievirus B1.";			
RL	Virolgy 156:64-73(1987).			
CC	-!- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE PROTEASES.			
CC	-!- FUNCTION: IT IS THOUGHT THAT THE P2C PROTEIN ATTACHES TO VESICULAR MEMBRANES AND IS ASSOCIATED WITH VIRAL RNA SYNTHESIS.			
CC	-!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS, EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2, VP3, AND VP4.			
CC	-!- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS. CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/VP2 IS CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.			
CC	-!- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.			
CC	-!- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; M16560; AAC00531.1; -			
DR	PIR; A26353; GNNYB1.			
DR	HSSP; P03313; LCOV.			
DR	MEROPS; C03.011; -			
DR	MEROPS; C03.022; -			
DR	InterPro; IPR000199; Cys-protease-3c.			
DR	InterPro; IPR003138; Pico_P1A.			
DR	InterPro; IPR000081; Pico_P2A.			
DR	InterPro; IPR002527; Pico_P2B.			
DR	InterPro; IPR000605; RNA_helicase.			
DR	InterPro; IPR001205; RNA_pol_P3D.			
DR	InterPro; IPR001676; Rhv.			

P34703 caenorhabdi
P37800 toscana vir
P93329 medicago tr
P22925 petunia hyb
P14317 homo sapien
Q03053 c genome po
P95649 rhodobacter
P54461 bacillus su
P50171 mus musculu
P39767 rhodopseudo
P52348 human herpe
Q58224 methanococc

ALIGNMENTS

```
DR Pfam: PF00548; Cys-protease-3C; 1.
DR DR PF02226; Pico_PIA; 1.
DR DR PF00947; Pico_P2A; 1.
DR DR PF01552; Pico_P2B; 1.
DR DR PF00073; rhv; 3.
DR DR PF00680; RNA_dep_RNA_pol; 1.
DR DR PF00910; RNA_helicase; 1.
DR DR ProDom: PD001125; Cys-protease-3C; 1.
DR DR ProDom: PD001274; Pico_P2B; 1.
DR DR ProDom: PD001306; Pico_P2A; 1.
DR KW Polyprotein; Coat protein; Core protein; Transferase; Myristate;
RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT CHAIN 2 69
FT CHAIN 70 332
FT CHAIN 333 570
FT CHAIN 571 848
FT CHAIN 849 998
FT CHAIN 999 1097
FT CHAIN 1098 1426
FT CHAIN 1427 1515
FT CHAIN 1516 1537
FT CHAIN 1538 1720
FT CHAIN 1721 2182
FT CHAIN 2182 1684
FT ACT_SITE 1684 1698
FT ACT_SITE 1698 1698
FT SEQUENCE 2182 AA; 243945 MW; 387B9391275859B1 CRC64;

Query Match 42.4%; Score 44.5; DB 1; Length 2182;
Best Local Similarity 28.6%; Pred. No. 37;
Matches 10; Conservative 1; Mismatches 5; Indels 19; Gaps 1;

QY 2 TTVGVFG-----IEQDWDRCV 17
|||
DB 846 TTGAFGQCGAGVYGVNVRVNRHLATREDWQRCV 880

RESULT 2
SOLC SCHJA STANDARD; PRT; 171 AA.
AC Q94743;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SORCIN
OS Schistosoma japonicum (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminths; Turbellarian Platyhelminths;
OC Rhabditophora; Euteleostomi; Neodermata; Medusozoa;
OC Neodermata; Trematoda; Digenea; Strigoida; Schistosomatoidea;
OC Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHINESE;
RA Clough K.A., Brindley P.J.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS, BELONGS
CC TO THE SORCIN/GRANULIN/CALPAIN LIGHT CHAIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: A39069; AAB17908.1;
CC HSP; P02593; IAKB.
CC InterPro: IPR002048; EF-hand.
CC Pfam: PF00036; efhand; 4.
CC SMART: SM00054; EFh; 3.

DR PROSITE; PS00018; EF_HAND; 3.
DR KW Calcium-binding; Repeat.
DR FT CA_BIND 16 27 EF-HAND 1 (POTENTIAL).
DR FT CA_BIND 53 64 EF-HAND 2 (POTENTIAL).
DR FT CA_BIND 83 94 EF-HAND 3 (POTENTIAL).
DR SQ SEQUENCE 171 AA; 19687 MW; 251FB948FA07ADDD CRC64;

Query Match 40.0%; Score 42; DB 1; Length 171;
Best Local Similarity 43.8%; Pred. No. 6.8;
Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCV 16
|||
DB 63 NEFLGLFKYVDWQTC 78

RESULT 3
DPY4_MOUSE STANDARD; PRT; 572 AA.
ID DPY4_MOUSE
AC O35098; O08886;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DIHYDROPYRIMIDINASE RELATED PROTEIN-4 (DRP-4) (ULIP4 PROTEIN).
GN DPYSL4 OR ULIP4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Hamajima N., Kato Y., Kouwaki M., Wada Y., Sasaki M., Nonaka M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=98314496; PubMed=9652388;
RA Byk T., Ozon S., Sobel A.;
RT "The Ulip family phosphoproteins -- common and specific properties.";
RL Eur. J. Biochem. 254:14-24(1998).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DIHYDROPYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: A8006715; BAA21888.1;
CC MGD; Y09079; CAA70299.1;
CC InterPro: IPR002195; Dihydroorotase.
CC Pfam: PF00744; Dihydroorotase; 1.
CC CONFLICT 125 126 ER -> DG (IN REF. 2).
CC CONFLICT 354 354 G -> V (IN REF. 2).
CC CONFLICT 420 420 F -> I (IN REF. 2).
CC SQ SEQUENCE 572 AA; 61961 MW; 37671129FC02C7AF CRC64;

Query Match 39.0%; Score 41; DB 1; Length 572;
Best Local Similarity 47.4%; Pred. No. 35;
Matches 9; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 5 GVFGIEQDWDRCVIS 19
|||
DB 354 GVNGIERNSVWWEKCVAS 372
```


Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 TTGVFGIEQDWDRCVIS 19
:| | | | :| | | |
Db 343 STVTVLGITSDFYQCSIS 360

```
RESULT 6
VIT_ANTGR
ID VIT_ANTGR STANDARD; PRT; 1790 AA.
AC Q05808;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VITELLOGENIN PRECURSOR [CONTAINS: YP47; YP160].
GN VITG.
OS Anthonomus grandis (Boll weevil).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Phytophaga; Curculionidae; Curculioninae; Anthonomini;
OC Anthonomus.
OX NCBI_TaxID=7044;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92277665; PubMed=1593641;
RA Trewitt P.M., Heilmann L.J., Degrugillier S.S., Kumaran A.K.;
RT "The boll weevil vitellogenin gene: nucleotide sequence, structure,
RT and evolutionary relationship to nematode and vertebrate vitellogenin
RT genes."
RL J. Mol. Evol. 34:478-492(1992).
CC -!- FUNCTION: PRECURSOR OF THE EGG-YOLK PROTEINS THAT ARE SOURCES OF
CC NUTRIENTS DURING EMBRYONIC DEVELOPMENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M72980; AAA27740.1;
CC InterPro: IPR001747; Vitellogenin_N.
CC InterPro: IPR001846; Vwd.
CC Pfam; PF01347; Vitellogenin_N; 1.
CC SMART; SM00094; vwd; 1.
CC SMART; SM00216; VWD; 1.
CC Glycoprotein; Storage protein; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1790 VITELLOGENIN.
FT CHAIN ? YP47.
FT CHAIN ? 1790 YP160.
FT CARBOHYD 219 219 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 554 554 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 893 893 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1345 1345 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1416 1416 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1430 1430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1480 1480 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1699 1699 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1735 1735 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 306 309 POLY-GLN.
FT DOMAIN 381 385 POLY-GLN.
FT DOMAIN 1039 1042 POLY-GLN.
SQ SEQUENCE 1790 AA; 205858 MW; E0B07A5B1E70D0DA CRC64;
```

Query Match 38.6%; Score 40.5; DB 1; Length 1790;
Best Local Similarity 61.5%; Pred. No. 1.4e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 5 GVFGI-EQDWDRC 16
| | | | | | | | | |
Db 1122 GVAGVWEQLWDKC 1134

```
RESULT 7
TXAL_ANESU
ID TXAL_ANESU STANDARD; PRT; 46 AA.
AC P01533;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE NEUROTOXIN 1 (TOXIN ATX-1).
OS Anemonia sulcata (Snake-locks sea anemone).
OC Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Actiniaria;
OC Nynanthacea; Actiniidae; Anemonia.
OX NCBI_TaxID=6108;
RN [1]
RP SEQUENCE.
RX MEDLINE=79024590; PubMed=29753;
RA Wunderer G., Eulitz M.;
RT "Amino-acid sequence of toxin I from Anemonia sulcata."
RL Eur. J. Biochem. 89:11-17(1978).
RN [2]
RP STRUCTURE BY NMR.
RX MEDLINE=88111667; PubMed=2892680;
RA Widmer H., Wagner G., Schweitz H., Lazdunski M., Wuethrich K.;
RT "The secondary structure of the toxin ATX Ia from Anemonia sulcata in
RT aqueous solution determined on the basis of complete
RT sequence-specific 1H-NMR assignments."
RL Eur. J. Biochem. 171:177-192(1988).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE=90160299; PubMed=2576133;
RA Widmer H., Billster M., Wuethrich K.;
RT "Three-dimensional structure of the neurotoxin ATX Ia from Anemonia
RT sulcata in aqueous solution determined by nuclear magnetic resonance
RT spectroscopy."
RL Proteins 6:357-371(1989).
CC -!- FUNCTION: BINDS SPECIFICALLY TO THE SODIUM CHANNEL, THEREBY
CC DELAYING ITS INACTIVATION DURING SIGNAL TRANSDUCTION. THUS
CC IT STRONGLY STIMULATES MAMMALIAN CARDIAC MUSCLE CONTRACTION.
CC -!- SIMILARITY: TO OTHER SEA ANEMONE SODIUM CHANNEL INHIBITORY
CC TOXINS. THIS IS A TYPE I TOXIN.
CC PIR; A01796; TZA21.
CC DR PDB; 1ATX; 15-JUL-91.
CC DR InterPro; IPR000693; Anemome_toxin.
CC DR Pfam; PF00706; toxin_4; 1.
CC KW Toxin; Sodium channel inhibitor; Multigene family; 3D-structure.
FT DISULFID 4 43
FT DISULFID 6 34
FT DISULFID 27 44
FT VARIANT 3 3 A -> P (ABOUT 20% OF THE MOLECULES).
FT STRAND 3 5
FT TURN 7 8
FT TURN 14 15
FT STRAND 18 23
FT TURN 29 30
FT STRAND 32 34
FT STRAND 42 45
SQ SEQUENCE 46 AA; 4814 MW; 862C1E21FDA5432D CRC64;
```

Query Match 38.1%; Score 40; DB 1; Length 46;
Best Local Similarity 42.9%; Pred. No. 3.7;
Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 TVGVFGIEQDWDRC 16
| | | | | | | | | |
Db 21 TIWVFGCPGWNKC 34

RESULT 8

```

FER_THEAC
ID  FER_THEAC      STANDARD;      PRT;      142 AA.
AC  P00218; Q9V304;
DT  21-JUL-1986 (Rel. 01, Created)
DT  21-JUL-1986 (Rel. 01, Last sequence update)
DT  20-AUG-2001 (Rel. 40, Last annotation update)
DE  ZINC-CONTAINING FERREDOXIN.
GN  ZFX OR TAI448.
OS  Thermoplasma acidophilum.
OC  Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
OC  Thermoplasma.
OX  NCBI_TaxID=2303;
RN  [1]
RP  SEQUENCE.
RC  STRAIN=DSM 1728;
RA  Wakabayashi S., Fujimoto N., Wada K., Matsubara H., Kerscher L.,
RA  Oesterhelmt D.;
RT  "Amino acid sequence of a ferredoxin from thermoacidophilic
RL  archaeobacteria, Thermoplasma acidophilum.";
RN  FEBS Lett. 162:21-24(1983).
[2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=DSM 1728;
RA  MEDLINE=20479972; PubMed=11029001;
RA  Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA  Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT  "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RL  acidophilum.";
RL  Nature 407:508-513(2000).
[3]
RN  SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC  STRAIN=HO-62;
RX  MEDLINE=99367440; PubMed=10438486;
RA  Cosper N.J., Stalhandske C.M.V., Iwasaki H., Oshima T., Scott R.A.,
RA  Iwasaki T.;
RT  "Structural conservation of the isolated zinc site in archaeal zinc-
RT  containing ferredoxins as revealed by X-ray absorption spectroscopic
RT  analysis and its evolutionary implications.";
RN  J. Biol. Chem. 274:23160-23168(1999).
[4]
RN  SEQUENCE OF 1-15, AND CHARACTERIZATION.
RC  STRAIN=HO-62;
RX  MEDLINE=97166191; PubMed=9013590;
RA  Iwasaki T., Suzuki T., Kon T., Imai T., Urushiyama A., Ohmori D.,
RA  Oshima T.;
RT  "Novel zinc-containing ferredoxin family in thermoacidophilic
RL  archaea.";
RL  J. Biol. Chem. 272:3453-3458(1997).
CC  -!- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
CC  ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC  -!- COFACTOR: BINDS ONE 3FE-4S AND ONE 4FE-4S CLUSTER AND ONE ZINC
CC  ATOM.
CC  -!- MASS SPECTROMETRY: MW=15961; MW_ERR=10; METHOD=MALDI.
CC  -!- SIMILARITY: BELONGS TO THE BACTERIAL TYPE FERREDOXIN FAMILY.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL; AL445067; CAC12568.1; -
DR  EMBL; AB023294; BA82797.1; -
DR  PIR; A00222; FEYTA.
DR  HSSP; P55907; 1XER.
DR  InterPro; IPR001450; 4FE4S_ferredoxin.
DR  Pfam; PF00037; fer4. 2.
DR  PROSITE; PS00198; 4FE4S-FERREDOXIN; 1.
KW  Electron transport; Iron-sulfur; Repeat; 4Fe-4S; 3Fe-4S; Zinc;
Complete proteome.
FT  INIT_MET 0

```

```

FT  DOMAIN      12 59
FT  DOMAIN      60 142
FT  METAL        30
FT  METAL        33
FT  METAL        33
FT  METAL        57
FT  METAL       116
FT  METAL       116
FT  METAL       68
FT  METAL       71
FT  METAL       74
FT  METAL       78
FT  METAL       78
FT  METAL      123
FT  METAL      126
FT  METAL      129
FT  METAL      133
FT  METAL      101
FT  VARIANT      101
FT  VARIANT      105
FT  SEQUENCE    142 AA; 15962 MW; 0E1970A7FBD407D5 CRC64;

Query Match      38.1%; Score 40; DB 1; Length 142;
Best Local Similarity 38.9%; Pred. No. 12;
Matches 7; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

QY  2 TTVGVFG--IEQWDRCV 17
Db  52 TKLGIHTHVAVDWDCCI 69

RESULT 9
FLGE_TREPH
ID  FLGE_TREPH  STANDARD;      PRT;      463 AA.
AC  Q56326;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  01-NOV-1997 (Rel. 35, Last annotation update)
DE  FLAGELLAR HOOK PROTEIN FLGE.
GN  FLGE.
OS  Treponema phagedenis.
OC  Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX  NCBI_TaxID=162;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=KAZAN 5;
RX  MEDLINE=94266716; PubMed=8206841;
RA  Limberger R.J., Slivinski L.L., Samsonoff W.A.;
RT  "Genetic and biochemical analysis of the flagellar hook of Treponema
RL  phagedenis.";
RL  J. Bacteriol. 176:3631-3637(1994).
CC  -!- SIMILARITY: BELONGS TO THE FLAGELLA BASAL BODY ROD PROTEINS
CC  FAMILY.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL; U04619; AAA73467.1; -
DR  InterPro; IPR001444; Flag_bb_rod.
DR  Pfam; PF00460; flg_bb_rod; 1.
DR  PROSITE; PS00588; FLAGELLA_BB_ROD; 1.
KW  Flagella.
SQ  SEQUENCE 463 AA; 49527 MW; 960548C364CA35BC CRC64;

Query Match      38.1%; Score 40; DB 1; Length 463;
Best Local Similarity 54.5%; Pred. No. 41;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY  2 TTVGVFGIEQD 12
    || | ||:::|

```


isoenzymes, and cloning and expression of a full-length cDNA.";
RL Biochem. J. 327:45-49(1997).
[2]

RP SEQUENCE FROM N.A.

RA Stinchi S., Stirling J.L., Orlacchio A., Beccari T.;
RT "Promoter characterization and structure of the gene encoding mouse
RT lysosomal alpha-mannosidase";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
[3]

RP SEQUENCE OF 3-1012 FROM N.A.

RC TISSUE=Liver;

RA MEDLINE-97449314; PubMed-9305783;

RA Merkle R.K., Zhang Y., Ruest P.J., Lal A., Liao Y.-F., Moremen K.W.;

RT "Cloning, expression, purification, and characterization of the murine
RT lysosomal acid alpha-mannosidase";

RL Biochim. Biophys. Acta 1336:132-146(1997).

CC -!- FUNCTION: NECESSARY FOR THE CATABOLISM OF N-LINKED CARBOHYDRATES
RELEASED DURING GLYCOPROTEIN TURNOVER.

CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING
ALPHA-D-MANNOSE RESIDUES IN ALPHA-D-MANNOSIDES.

CC -!- SUBCELLULAR LOCATION: LYSOSOMAL.

CC -!- SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U29947; AAC53369.1; ALT_INIT.

DR EMBL; AF044192; AAC78560.1; .

DR EMBL; AF044174; AAC78560.1; JOINED.

DR EMBL; AF044175; AAC78560.1; JOINED.

DR EMBL; AF044176; AAC78560.1; JOINED.

DR EMBL; AF044177; AAC78560.1; JOINED.

DR EMBL; AF044178; AAC78560.1; JOINED.

DR EMBL; AF044179; AAC78560.1; JOINED.

DR EMBL; AF044180; AAC78560.1; JOINED.

DR EMBL; AF044181; AAC78560.1; JOINED.

DR EMBL; AF044182; AAC78560.1; JOINED.

DR EMBL; AF044183; AAC78560.1; JOINED.

DR EMBL; AF044184; AAC78560.1; JOINED.

DR EMBL; AF044185; AAC78560.1; JOINED.

DR EMBL; AF044186; AAC78560.1; JOINED.

DR EMBL; AF044187; AAC78560.1; JOINED.

DR EMBL; AF044188; AAC78560.1; JOINED.

DR EMBL; AF044189; AAC78560.1; JOINED.

DR EMBL; AF044190; AAC78560.1; JOINED.

DR EMBL; AF044191; AAC78560.1; JOINED.

DR EMBL; U87240; AAC09470.1; ALT_INIT.

DR MGB; MGI:107286; Man2b1.

DR InterPro; IPR000602; Glyco_hydro_38.

DR Pfam; PF01074; Glyco_hydro_38; 1.

KW Hydrolase; Glycosidase; Glycoprotein; Lysosome; Zymogen; Signal.

FT SIGNAL 1 43

FT CHAIN 44 1012

FT CARBOHYD 309 309

FT CARBOHYD 344 344

FT CARBOHYD 366 366

FT CARBOHYD 488 488

FT CARBOHYD 496 496

FT CARBOHYD 543 543

FT CARBOHYD 632 632

FT CARBOHYD 645 645

FT CARBOHYD 692 692

FT CARBOHYD 766 766

FT CARBOHYD 930 930

FT CONFLICT 45 48

FT LASG -> PGARA (IN REF. 3).

SEQUENCE 1012 AA; 114480 MW; 4FE9E5AED74709C CRC64;

Query Match 38.1%; Score 40; DB 1; Length 1012;
Best Local Similarity 47.1%; Pred. No. 92;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 TVGVFGIEQDWDRCVIS 19

DB 715 TVGPIPVDDMGKEVIS 731

||||| : ||| : |||

RESULT 13

NRTB_PHOLA

ID NRTB_PHOLA STANDARD; PRT; 279 AA.

AC Q51881;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE NITRATE TRANSPORT PERMEASE PROTEIN NRTB.

GN NRTB.

OS Phormidium lamosum.

OC Bacteria; Cyanobacteria; Oscillatoriales; Phormidium.

OX NCBI_TaxID=32059;

RP [1]

SEQUENCE FROM N.A.

CC STRAIN=OH-1-P-CL1;

CC MEDLINE=95375238; PubMed=7647306;

CC Merchan F., Kindle K.L., Llama M.J., Serra J.L., Fernandez E.;

RT "Cloning and sequencing of the nitrate transport system from the
RT thermophilic, filamentous cyanobacterium Phormidium lamosum:
RT comparative analysis with the homologous system from Synechococcus
sp. PCC 7942";

RT Plant Mol. Biol. 28:759-766(1995).

CC -!- FUNCTION: PROBABLY PART OF A HIGH-AFFINITY BINDING-PROTEIN-
DEPENDENT TRANSPORT SYSTEM FOR NITRATE. PROBABLY RESPONSIBLE FOR
THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE (BY
SIMILARITY).

CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC -!- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE CYST
SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Z19598; CAA79657.1; .

DR InterPro; IPR000515; BPD_transp.

DR Pfam; PF00528; BPD_transp; 1.

DR PROSITE; PS00402; BPD_TRANSPP_INN_MEMBER; FALSE_NEG.

KW Transport; Transmembrane; Nitrate assimilation.

FT TRANSMEM 30 50

FT TRANSMEM 98 118

FT TRANSMEM 151 171

FT TRANSMEM 196 216

FT TRANSMEM 217 237

FT TRANSMEM 249 269

FT TRANSMEM 279 30204

SEQUENCE 279 AA; 30204 MW; 5BB86C4B02EFC5F6 CRC64;

Query Match 37.1%; Score 39; DB 1; Length 279;

Best Local Similarity 57.1%; Pred. No. 35;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWD 14

DB 166 NTAAGVQNPQDYN 179

||||| : ||| : |||

RESULT 14

MLTB_ECOLI

DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN AF0028.
GN AF0028
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -!- SIMILARITY: STRONG. TO M.JANNASCHII MJ1678, M.THERMOAUTOTROPHICUM
CC MTH1067 AND A.FULGIDUS AF0181.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001105; AAB91200.1; -.
DR TIGR; AF0028; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 371 AA; 41930 MW; B22056D00437BEAA CRC64;

Query Match 37.1%; Score 39; DB 1; Length 371;
Best Local Similarity 33.3%; Pred. No. 48;
Matches 5; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 4 VGVFGIEQDWDRCVI 18
Db :|||:|:|:|:
Db 217 IGVGMGKEEWAIVL 231

Search completed: March 14, 2002, 09:24:52
Job time: 908 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:10:44 ; Search time 56.96 seconds
(without alignments)
25.409 Million cell updates/sec

Title: US-09-765-739A-5

Perfect score: 105
Sequence: 1 NTTGVFGIEQDWDRCVIS 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : PIR_58:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	286	2 JE0219	28k surface antige
2	77	73.3	280	2 JE0217	28k surface antige
3	73	69.5	276	2 JE0218	28k surface antige
4	51	48.6	278	2 JE0216	28k surface antige
5	47	44.8	160	2 A75466	2-demethylmenaquin
6	46	43.8	284	2 I40882	major antigenic pr
7	44.5	42.4	2182	1 GNNYB1	genome polyprotein
8	43	41.0	5188	2 B85547	probable RTX famil
9	42.5	40.5	417	2 E84430	probable inositol
10	42	40.0	244	2 E84885	hypothetical prote
11	41	39.0	142	2 F83359	hypothetical prote
12	41	39.0	299	2 S68198	probable transcrip
13	41	39.0	573	2 T07948	mandelonitrile lya
14	41	39.0	576	2 T08073	mandelonitrile lya
15	41	39.0	688	2 JC2248	polyprotein - swee
16	40.5	38.6	205	2 A72479	hypothetical prote
17	40.5	38.6	1790	1 S27772	vitellogenin precu
18	40	38.1	46	1 TZA21	toxin I - snake-lo
19	40	38.1	143	1 FEYTA	ferredoxin [3Fe-4S
20	40	38.1	349	2 F70357	lipoprotein - Aqu
21	40	38.1	361	2 C85430	MAP kinase like pr
22	40	38.1	398	2 C83499	probable transglyc
23	40	38.1	436	2 T28066	hypothetical prote
24	40	38.1	530	2 T23255	hypothetical prote
25	40	38.1	540	2 S63299	sugar transport pr
26	40	38.1	1012	2 T42385	alpha-mannosidase
27	40	38.1	1090	2 C86577	pp2-transglycolas
28	40	38.1	1090	2 D72048	pp2-transglycolas
29	39.5	37.6	116	2 T36999	probable transposa

30 39 37.1 216 2 B85712 unknown protein en
31 39 37.1 265 2 T43123 probable positive
32 39 37.1 279 1 S56642 nitrate transport
33 39 37.1 331 2 A70884 probable ribF prot
34 39 37.1 331 2 T44902 probable riboflavi
35 39 37.1 361 2 A65050 membrane-bound lyl
36 39 37.1 361 2 B85918 membrane-bound lyl
37 39 37.1 367 2 H83088 membrane-bound lyl
38 39 37.1 371 2 D69253 conserved hypoteth
39 39 37.1 377 2 F69008 acetyltransferase
40 39 37.1 397 2 I58099 gene P2X3 protein
41 39 37.1 397 2 S60334 purinoceptor P2X -
42 39 37.1 462 2 D84751 hypothetical prote
43 39 37.1 503 2 T40997 probable short-cha
44 39 37.1 548 2 E72457 probable phenylala
45 39 37.1 574 2 T50766 mandelonitrile lya

ALIGNMENTS

RESULT
JE0219
28k surface antigen 2 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0219
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrige, M.J.; Alleman, A.R
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0219
A:Molecule type: DNA
A:Residues: 1-286 <RED>
A:Cross-references: GB:AF062761

Query Match 100.0%; Score 105; DB 2; Length 286;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NTTGVFGIEQDWDRCVIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78
|||||
|||||

RESULT 2
JE0217
28k surface antigen 4 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0217
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrige, M.J.; Alleman, A.R
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0217
A:Molecule type: DNA
A:Residues: 1-280 <RED>
A:Cross-references: GB:AF062761

Query Match 73.3%; Score 77; DB 2; Length 280;
Best Local Similarity 68.4%; Pred. No. 4.4e-05;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NTTGVFGIEQDWDRCVIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78
|||||
|||||

RESULT 3

28k surface antigen 5 - Ehrlichia chaffensis
 N:Alternate names: MAP1
 C:Species: Ehrlichia chaffensis
 C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
 C:Accession: JE0218
 R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrige, M.J.; Alleman, A.R.
 Biochem. Biophys. Res. Commun. 247, 636-643, 1998
 A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
 A:Reference number: JE0216; MUID:98321180
 A:Accession: JE0218
 A:Molecule type: DNA
 A:Residues: 1-276 <RED>
 A:Cross-references: GB:AF062761

Query Match 69.5%; Score 73; DB 2; Length 276;
 Best Local Similarity 68.4%; Pred. No. 0.0002;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEDWDRCVIS 19
 |||||:::||||
 Db 59 NTTGVFGELKQNDGSAIS 77

RESULT 4

28k surface antigen 3 - Ehrlichia chaffensis
 N:Alternate names: MAP1
 C:Species: Ehrlichia chaffensis
 C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
 C:Accession: JE0216
 R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrige, M.J.; Alleman, A.R.
 Biochem. Biophys. Res. Commun. 247, 636-643, 1998
 A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
 A:Reference number: JE0216; MUID:98321180
 A:Accession: JE0216
 A:Molecule type: DNA
 A:Residues: 1-278 <RED>
 A:Cross-references: GB:AF062761

Query Match 48.6%; Score 51; DB 2; Length 278;
 Best Local Similarity 50.0%; Pred. No. 0.82;
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGIEDWD 14
 ||:::||||
 Db 60 NPTVALYGLKQDWE 73

RESULT 5

2-demethylmenaquinone 2-C-methyltransferase (EC 2.1.1.-) DR0859 [similarity] - Deinococcus
 C:Species: Deinococcus radiodurans
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
 C:Accession: A75466
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: A75466
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-160 <WHI>
 A:Cross-references: GB:AE001940; GB:AE000513; NID:96458577; PIDN:AAF10437.1; PID:96458577
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0859
 A:Map position: 1

C:Keywords: methyltransferase

Query Match 44.8%; Score 47; DB 2; Length 160;
 Best Local Similarity 46.7%; Pred. No. 2.1;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 4 VGVFGIEDWDRCVI 18
 :|||:::|:
 Db 79 LGVFGVNGWEGVII 93

RESULT 6

I40882
 major antigenic protein - heartwater rickettsia
 C:Species: Cowdria ruminantium (heartwater rickettsia)
 C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
 C:Accession: I40882; S42827
 R:van Vliet, A.H.; Jongejans, F.; van Kleef, M.; van der Zeijst, B.A.
 Infect. Immun. 62, 1451-1456, 1994
 A:Title: Molecular cloning, sequence analysis, and expression of the gene encoding th
 A:Reference number: I40882; MUID:94178956
 A:Accession: I40882
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-284 <RES>
 A:Cross-references: EMBL:X74250; NID:9454266; PIDN:CAA52309.1; PID:9454267
 C:Genetics:
 A:Gene: map1

Query Match 43.8%; Score 46; DB 2; Length 284;
 Best Local Similarity 58.3%; Pred. No. 5.5;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVGVFGIEDWD 14
 ||:::||||
 Db 63 TKAVFGLKQDWD 74

RESULT 7

GNNYB1
 genome polyprotein - coxsackievirus B1
 N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core
 polymerase (EC 2.7.7.48)
 C:Species: coxsackievirus B1
 C>Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
 C:Accession: A26353
 R:Iizuka, N.; Kuge, S.; Nomoto, A.
 Virology 156, 64-73, 1987
 A:Title: Complete nucleotide sequence of the genome of coxsackievirus B1.
 A:Reference number: A26353; MUID:87122157
 A:Accession: A26353
 A:Molecule type: genomic RNA
 A:Residues: 1-2182 <II?>
 A:Cross-references: GB:M16560; NID:G323417; PIDN:AAC00531.1; PID:G323418
 C:Superfamily: poliovirus genome polyprotein
 C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase
 F:1-69/Product: coat protein 1A #status predicted <PIA>
 F:70-332/Product: coat protein 1B #status predicted <PIB>
 F:333-570/Product: coat protein 1C #status predicted <PIC>
 F:571-848/Product: coat protein 1D #status predicted <PID>
 F:849-998/Product: core protein 2A #status predicted <C2A>
 F:999-1097/Product: core protein 2B #status predicted <C2B>
 F:1098-1426/Product: core protein 2C #status predicted <C2C>
 F:1427-1515/Product: protein 3A #status predicted <P3A>
 F:1516-1537/Product: genome-linked protein VPg #status predicted <VPG>
 F:1538-1720/Product: proteinase #status predicted <PTS>
 F:1721-2182/Product: RNA-directed RNA polymerase #status predicted <RNS>
 F:1518/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 42.4%; Score 44.5; DB 1; Length 2182;

```
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84885
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.B.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: E84885
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-244 <STO>
A:Cross-references: GB:AE002093; NID:g4895249; PIDN:AAD32834.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g45010
A:Map position: 2

Query Match 40.0%; Score 42; DB 2; Length 244;
Best Local Similarity 43.8%; Pred. No. 22;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TTGVGFGIEQDWDRC 16
Db 61 NMTTGIFGCAEDPESC 76

RESULT 11
F83359
hypothetical protein PA2282 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83359
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.: Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: F83359
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-142 <STO>
A:Cross-references: GB:AE004654; GB:AE004091; NID:g9948311; PIDN:AAG05670.1; GSPDB:GN
C:Genetics:
A:Gene: PA2282

Query Match 39.0%; Score 41; DB 2; Length 142;
Best Local Similarity 50.0%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 6 VFGIEQDWDRCV 17
Db 62 VEGLDHWDRLI 73

RESULT 12
S68198
probable transcription regulator (carotenoid gene cluster orf 11) - Myxococcus xanthu
C:Species: Myxococcus xanthus
C:Date: 24-Aug-1996 #sequence_revision 13-Mar-1997 #text_change 08-Oct-1999
C:Accession: S68198; S67958
R:Botella, J.; Murillo, F.; Ruiz-vazquez, R.
submitted to the EMBL Data Library, November 1994
A:Description: A cluster of structural and regulatory genes for light-induced caroten
A:Reference number: S68198
A:Accession: S68198
A:Molecule type: DNA
A:Residues: 1-299 <BOT>
A:Cross-references: EMBL:Z21955; NID:g577589; PIDN:CAA79965.1; PID:g577593
R:Botella, J.A.; Murillo, F.J.; Ruiz-Vazquez, R.
```

```
Best Local Similarity 28.6%; Pred. No. 79;
Matches 10; Conservative 1; Mismatches 5; Indels 19; Gaps 1;

Qy 2 TTGVGFG-----IEQDWDRCV 17
Db 846 TTGAFGQGSQGVYGVNRYVVRHRLATREDWQRCV 880

RESULT 8
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B85547
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85547
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:Z06
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0615

Query Match 41.0%; Score 43; DB 2; Length 5188;
Best Local Similarity 50.0%; Pred. No. 3.4e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 2 TTGVGFGIEQDWDRCV 17
Db 4803 TTSGVAAMDYDWDGAV 4818

RESULT 9
E84430
probable inositol polyphosphate-5-phosphatase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84430
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.B.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: E84430
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-417 <STO>
A:Cross-references: GB:AE002093; NID:g4522008; PIDN:AAD21781.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g01900
A:Map position: 2

Query Match 40.5%; Score 42.5; DB 2; Length 417;
Best Local Similarity 60.0%; Pred. No. 31;
Matches 9; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Qy 5 GVFGIEQDWDRCVIS 19
Db 144 GINGISQDF-RCIIS 157

RESULT 10
E84885
hypothetical protein At2g45010 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
```

Eur. J. Biochem. 233, 238-248, 1995
A:Title: A cluster of structural and regulatory genes for light-induced carotenogenesis
A:Reference number: S67950; MUID:96061955
A:Accession: S67958
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 7-68 <BOW>
A:Cross-references: EMBL:Z21955
C:Keywords: DNA binding

Query Match 39.0%; Score 41; DB 2; Length 299;
Best Local Similarity 66.7%; Pred No. 39;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 4 VGVFGIEQDWR 15
DB 284 VHVFGQEDWR 295

RESULT 13
T07948
mandelonitrile lyase (EC 4.1.2.10) 3 - black cherry
C:Species: Prunus serotina (black cherry)
C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 17-Mar-2000
C:Accession: T07948
R:Hu, Z.; Poulton, J.E.
submitted to the EMBL Data Library, July 1997
A:Description: Prunus serotina (R)-(+)-mandelonitrile lyase isoform MDL3 precursor, mRNA
A:Reference number: Z16239
A:Accession: T07948
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-573 <HUZ>
A:Cross-references: EMBL:AF013161; NID:g2343180; PIDN:AAB67714.1; PID:g2343181
C:Gene: MDL3
C:Function:
A:Description: catalyzes dissociation of (R)-mandelonitrile to hydrogen cyanide and benzaldehyde
C:Superfamily: alcohol oxidase
C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 39.0%; Score 41; DB 2; Length 573;
Best Local Similarity 50.0%; Pred No. 76;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 2 TTVGVFGIEQDWR 19
DB 342 STVTVLGITSDFYQCSIS 359

RESULT 14
T08073
mandelonitrile lyase (EC 4.1.2.10) 2 precursor - black cherry
C:Species: Prunus serotina (black cherry)
C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 17-Mar-2000
C:Accession: T08073
R:Hu, Z.; Poulton, J.E.
submitted to the EMBL Data Library, December 1997
A:Reference number: Z16335
A:Accession: T08073
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-576 <HUZ>
A:Cross-references: EMBL:AF040078; NID:g2773273; PIDN:AAB96763.1; PID:g2773274
C:Function:
A:Description: catalyzes the dissociation of (S)-(-)-mandelonitrile to benzaldehyde and
A:Pathway: hydrolysis of cyanogenic disaccharides
C:Superfamily: alcohol oxidase
C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 39.0%; Score 41; DB 2; Length 576;
Best Local Similarity 50.0%; Pred No. 76;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 2 TTVGVFGIEQDWR 19
DB 343 STVTVLGITSDFYQCSIS 360

RESULT 15
JC2248
polyprotein - sweet potato feathery mottle virus (fragment)
N:Contains: coat protein; nuclear inclusion b protein
C:Species: sweet potato feathery mottle virus
C:Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 17-Nov-2000
A:Accession: JC2248; PC2164
R:Mori, M.; Usugi, T.; Hayashi, T.; Nishiguchi, M.
Biosci. Biotechnol. Biochem. 58, 965-967, 1994
A:Title: Nucleotide sequence at the 3'-terminal region of sweet potato feathery mottle virus polyprotein
A:Reference number: JC2248; MUID:94289871
A:Accession: JC2248
A:Molecule type: mRNA
A:Residues: 1-688 <MORI>
A:Cross-references: DDBJ:D16664; NID:g391906; PIDN:BAA04068.1; PID:d1004586; PID:g473
A:Experimental source: clone PMO-2.3
A:Note: the authors translated the codon CAT for residue 170 as Asp, CAT for residue 171 as Asp
A:Accession: PC2164
A:Molecule type: protein
A:Residues: 453-462 <MOR2>
C:Superfamily: tobacco etch virus genome polyprotein
C:Keywords: coat protein; polyprotein
F:1-373/Product: nuclear inclusion b protein (fragment) #status predicted <NIB>
F:374-688/Product: coat protein #status predicted <COP>
F:373-374/Cleavage site: Gln-Ser (viral proteinase) #status predicted

Query Match 39.0%; Score 41; DB 2; Length 688;
Best Local Similarity 47.1%; Pred No. 91;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 3 TVGVFGIEQDWR 19
DB 73 TVGTFYQGDRLTS 89

Search completed: March 14, 2002, 09:10:45
Job time: 361 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:07:40 ; Search time 54.54 Seconds
(without alignments)
7.839 Million cell updates/sec

Title: US-09-765-739A-5
Perfect score: 105
Sequence: 1 NTTGVFGIEQDWDRCVIS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	286	4	US-08-953-326-15
2	77	73.3	280	4	US-08-953-326-17
3	73	69.5	276	4	US-08-953-326-18
4	73	69.5	280	3	US-08-733-230-4
5	73	69.5	280	4	US-08-953-326-4
6	51	48.6	278	4	US-08-953-326-16
7	44	41.9	287	3	US-08-733-230-2
8	44	41.9	287	4	US-08-953-326-2
9	44	41.9	898	2	US-08-808-982-5
10	44	41.9	898	4	US-09-306-902A-5
11	39	37.1	397	2	US-08-750-134A-9
12	39	37.1	397	4	US-09-363-745-9
13	39	37.1	397	4	US-09-191-136-16
14	39	37.1	397	4	US-09-191-136-17
15	39	37.1	943	2	US-08-808-982-7
16	39	37.1	943	4	US-09-306-902A-7
17	39	37.1	1018	1	US-08-072-610-2
18	39	37.1	1018	2	US-08-719-822B-2
19	39	37.1	1018	4	US-09-092-458-2
20	37.5	35.7	486	2	US-08-942-423-3
21	37	35.2	236	5	PCT-US95-04801-8
22	37	35.2	276	1	US-08-431-387-7
23	37	35.2	341	4	US-09-347-803-16
24	37	35.2	371	3	US-08-586-165-3
25	37	35.2	372	3	US-08-586-165-5
26	37	35.2	388	2	US-08-742-621-1
27	37	35.2	388	4	US-09-191-608-22

28	37	35.2	486	3	US-08-746-559A-5	Sequence 5, Appli
29	37	35.2	516	3	US-08-746-559A-4	Sequence 4, Appli
30	37	35.2	560	3	US-08-851-843A-6	Sequence 6, Appli
31	37	35.2	560	4	US-08-854-050-6	Sequence 6, Appli
32	37	35.2	1367	2	US-08-249-687C-2	Sequence 2, Appli
33	37	35.2	1367	2	US-08-625-819-2	Sequence 2, Appli
34	37	35.2	1367	3	US-08-746-559A-2	Sequence 2, Appli
35	36.5	34.8	546	2	US-08-942-423-4	Sequence 4, Appli
36	36	34.3	159	2	US-08-844-086-4	Sequence 4, Appli
37	36	34.3	159	3	US-09-018-211-4	Sequence 4, Appli
38	36	34.3	245	2	US-08-715-204-3	Sequence 3, Appli
39	36	34.3	245	3	US-09-162-597-3	Sequence 3, Appli
40	36	34.3	255	4	US-08-815-225-4	Sequence 4, Appli
41	36	34.3	833	2	US-08-844-086-2	Sequence 2, Appli
42	36	34.3	833	3	US-09-018-211-2	Sequence 2, Appli
43	36	34.3	877	2	US-08-907-166-8	Sequence 8, Appli
44	36	34.3	2285	4	US-09-308-375-2	Sequence 2, Appli
45	35.5	33.8	397	1	US-08-290-448A-76	Sequence 76, Appli

ALIGNMENTS

```

RESULT --1
US-08-953-326-15
; Sequence 15, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-15

```

Query Match 100.0%; Score 105; DB 4; Length 286;
Best Local Similarity 100.0%; Pred. No. 2.1e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 NTTGVFGIEQDWDRCVIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78

```

```

RESULT 2
US-08-953-326-17
; Sequence 17, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.

```


; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4

Query Match 69.5%; Score 73; DB 4; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.00028;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
||| ||||| :||| ||
Db 60 NTTGVFGLKQNDGSAIS 78

RESULT 6
US-08-953-326-16
; Sequence 16, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-16

Query Match 48.6%; Score 51; DB 4; Length 278;
Best Local Similarity 50.0%; Pred. No. 0.92;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWD 14
||| :||| :|||
Db 60 NPTVALYGLKQDWE 73

RESULT 7
US-08-733-230-2
; Sequence 2, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1

; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-2

Query Match 41.9%; Score 44; DB 3; Length 287;
Best Local Similarity 58.3%; Pred. No. 13;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVGVFGIEQDWD 14
| ||| :|||
Db 63 TOTVFLKKDWD 74

RESULT 8
US-08-953-326-2
; Sequence 2, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
US-08-953-326-2

Query Match 41.9%; Score 44; DB 4; Length 287;
Best Local Similarity 58.3%; Pred. No. 13;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVGVFGIEQDWD 14

Db 63 TQTVGLKRDWD 74

RESULT 9

US-08-808-982-5

Sequence 5, Application US/08808982

Patent No. 5939271

GENERAL INFORMATION:

APPLICANT: Tessier-Lavigne, Marc

APPLICANT: Leonardo, E. David

APPLICANT: Hink, Lindsay

APPLICANT: Masu, Masayuki

APPLICANT: Kazuko, Keino-Masu

TITLE OF INVENTION: Netrin Receptors

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

STREET: 268 BUSH STREET, SUITE 3200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/808,982

FILING DATE: 07-May-1999

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, RICHARD A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: UC96-217

TELEPHONE: (415) 343-4341

TELEFAX: (415) 343-4342

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 898 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-306-902A-5

Query Match 41.9%; Score 44; DB 2; Length 898;

Best Local Similarity 58.3%; Pred. No. 45;

Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 6 VFGIEQDWDRVC 17

DB 116 VFGLEEWQCVC 127

RESULT 10

US-08-808-982-5

Sequence 5, Application US/09306902A

Patent No. 6277585

GENERAL INFORMATION:

APPLICANT: Tessier-Lavigne, Marc

APPLICANT: Leonardo, E. David

APPLICANT: Hink, Lindsay

APPLICANT: Masu, Masayuki

APPLICANT: Kazuko, Keino-Masu

TITLE OF INVENTION: Netrin Receptors

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP

STREET: 268 BUSH STREET, SUITE 3200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/306,902A

FILING DATE: 07-May-1999

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: CRAWFORD, ARTHUR C.

REGISTRATION NUMBER: 25,327

REFERENCE/DOCKET NUMBER: 1430-116

TELEPHONE: (703) 816-4006

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 898 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 9:

US-09-306-902A-5

Query Match 41.9%; Score 44; DB 2; Length 898;

Best Local Similarity 58.3%; Pred. No. 45;

Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 6 VFGIEQDWDRVC 17

DB 116 VFGLEEWQCVC 127

RESULT 11

US-08-750-134A-9

Sequence 9, Application US/08750134A

Patent No. 5985603

GENERAL INFORMATION:

APPLICANT: VALERA, SOLEDAD

APPLICANT: BUELL, GARY

TITLE OF INVENTION: P2X RECEPTORS (PURINOCEPTOR FAMILY)

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/750,134A

FILING DATE: 22-JAN-1997

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: CRAWFORD, ARTHUR C.

REGISTRATION NUMBER: 25,327

REFERENCE/DOCKET NUMBER: 1430-116

TELEPHONE: (703) 816-4006

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 898 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 9:

US-09-306-902A-5

Query Match 41.9%; Score 44; DB 4; Length 898;

Best Local Similarity 58.3%; Pred. No. 45;

Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 6 VFGIEQDWDRVC 17

DB 116 VFGLEEWQCVC 127

; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-750-134A-9

Query Match 37.1%; Score 39; DB 2; Length 397;
Best Local Similarity 34.8%; Pred. No. 1.1e+02;
Matches 8; Conservative 4; Mismatches 3; Indels 8; Gaps 1;

QY 3 TVGVFGI-----EQDWDRCV 17
| | | | |
Db 235 TGGVLGKIGWVCDLKDWDQCI 257

RESULT 12

US-09-363-745-9
; Sequence 9, Application US/09363745
; Patent No. 6194162

; GENERAL INFORMATION:
; APPLICANT: VALERA, SOLEDAD
; APPLICANT: BUELL, GARY
; TITLE OF INVENTION: P2x RECEPTORS (PURINOCEPTOR FAMILY)
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/363,745
; FILING DATE:

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/750,134
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: CRAWFORD, ARTHUR C.
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 1430-116
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4006
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-363-745-9

Query Match 37.1%; Score 39; DB 4; Length 397;
Best Local Similarity 34.8%; Pred. No. 1.1e+02;
Matches 8; Conservative 4; Mismatches 3; Indels 8; Gaps 1;

QY 3 TVGVFGI-----EQDWDRCV 17
| | | | |
Db 235 TGGVLGKIGWVCDLKDWDQCI 257

RESULT 13

US-09-191-136-16
; Sequence 16, Application US/091911368

; Patent No. 6214581
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Lynch, Kevin J.
; APPLICANT: Burgard, Edward C.
; APPLICANT: Van Biesen, T.
; TITLE OF INVENTION: Nucleic Acids Encoding A Functional
; TITLE OF INVENTION: Human Purinoceptor P2x3 and P2X6 And Methods Of Production
; FILE REFERENCE: 6293 US.P1
; CURRENT APPLICATION NUMBER: US/09/191,136B
; CURRENT FILING DATE: 1998-11-13
; EARLIER APPLICATION NUMBER: US 09/008,526
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: US 09/008,185
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: US 60/071,298
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: US 60/071,669
; EARLIER FILING DATE: 1998-01-16
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-191-136-16

Query Match 37.1%; Score 39; DB 4; Length 397;
Best Local Similarity 34.8%; Pred. No. 1.1e+02;
Matches 8; Conservative 4; Mismatches 3; Indels 8; Gaps 1;

QY 3 TVGVFGI-----EQDWDRCV 17
| | | | |
Db 235 TGGVLGKIGWVCDLKDWDQCI 257

RESULT 14

US-09-191-136-17
; Sequence 17, Application US/091911368
; Patent No. 6214581
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Lynch, Kevin J.
; APPLICANT: Burgard, Edward C.
; APPLICANT: Van Biesen, T.
; TITLE OF INVENTION: Nucleic Acids Encoding A Functional
; TITLE OF INVENTION: Human Purinoceptor P2x3 and P2X6 And Methods Of Production
; FILE REFERENCE: 6293 US.P1
; CURRENT APPLICATION NUMBER: US/09/191,136B
; CURRENT FILING DATE: 1998-11-13
; EARLIER APPLICATION NUMBER: US 09/008,526
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: US 09/008,185
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: US 60/071,298
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: US 60/071,669
; EARLIER FILING DATE: 1998-01-16
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-191-136-17

Query Match 37.1%; Score 39; DB 4; Length 397;
Best Local Similarity 34.8%; Pred. No. 1.1e+02;
Matches 8; Conservative 4; Mismatches 3; Indels 8; Gaps 1;

QY 3 TVGVFGI-----EQDWDRCV 17
| | | | |
Db 235 TGGVLGKIGWVCDLDRWDQCI 257

RESULT 15
US-08-808-982-7
; Sequence 7, Application US/08808982
; Patent No. 5939271
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: Leonardo, E. David
; APPLICANT: Hink, Lindsay
; APPLICANT: Masu, Masayuki
; APPLICANT: Kazuko, Keino-Masu
; TITLE OF INVENTION: Netrin Receptors
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/808,982
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UC96-217
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
US-08-808-982-7

Query Match 37.1%; Score 39; DB 2; Length 943;
Best Local Similarity 50.0%; Pred. No. 3e+02; 3; Indels 0; Gaps 0;
Matches 6; Conservative 3; Mismatches 3

QY 6 VFGIEQDWDRCV 17
| | | | |
Db 118 LFGLEDYWCQCV 129

Search completed: March 14, 2002, 09:07:41
Job time: 842 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:09:37 ; Search time 111.55 Seconds
(without alignments)
12.617 Million cell updates/sec

Title: US-09-765-739A-5
Perfect score: 105
Sequence: 1 NTTVGVFGIEQDWDRCVIS 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	286	19 AAW51092	Ehrlichia chaffee
2	105	100.0	286	20 AAY06946	E. chafeensis OMP
3	105	100.0	286	21 AAB36186	Ehrlichia chaffee
4	105	100.0	286	22 AAU04196	Variable surface a
5	77	73.3	280	19 AAW51094	Ehrlichia chaffee
6	77	73.3	280	21 AAB36188	Ehrlichia chaffee
7	77	73.3	280	22 AAU04198	Variable surface a
8	74	70.5	280	20 AAY06948	E. chafeensis OMP
9	73	69.5	256	20 AAY06942	Ehrlichia chaffee
10	73	69.5	276	19 AAW51095	Ehrlichia chaffee
11	73	69.5	276	21 AAB36189	Ehrlichia chaffee

12	73	69.5	276	22	AAU04199	Variable surface a
13	73	69.5	280	19	AAW51089	Ehrlichia chaffee
14	73	69.5	280	21	AAB36183	Ehrlichia chaffee
15	73	69.5	280	22	AAU04193	Major antigenic pr
16	73	69.5	281	20	AAU06943	E. chafeensis OMP
17	71	67.6	288	20	AAU06959	E. canis p30 prote
18	62	59.0	280	20	AAU06962	Ehrlichia canis im
19	62	59.0	280	21	AAU06962	Ehrlichia canis im
20	58	55.2	278	21	AAU06961	E. canis p30-1 pro
21	58	55.2	278	20	AAU06961	E. canis p30-4 pro
22	54	51.4	276	20	AAU06964	Ehrlichia chaffee
23	51	48.6	278	19	AAW51093	E. chafeensis OMP
24	51	48.6	278	20	AAU06947	Ehrlichia chaffee
25	51	48.6	278	21	AAB36187	Variable surface a
26	51	48.6	278	22	AAU04197	E. chafeensis OMP
27	50	47.6	280	20	AAU06945	Cowdria ruminantii
28	44	41.9	287	19	AAW51088	Cowdria ruminantii
29	44	41.9	287	21	AAB36182	Major antigenic pr
30	44	41.9	287	22	AAU04192	Rat UNC-5 homologo
31	44	41.9	898	19	AAW78898	Arabidopsis thalia
32	42	40.0	198	21	AAG07319	Arabidopsis thalia
33	42	40.0	198	21	AAG38298	Arabidopsis thalia
34	42	40.0	220	21	AAG08266	Arabidopsis thalia
35	42	40.0	220	21	AAG14886	Arabidopsis thalia
36	42	40.0	220	21	AAG48818	Arabidopsis thalia
37	42	40.0	220	21	AAG48858	Arabidopsis thalia
38	42	40.0	244	21	AAG07318	Arabidopsis thalia
39	42	40.0	244	21	AAG38297	Arabidopsis thalia
40	41	39.0	521	15	AAW54988	Sweet potato feath
41	41	39.0	1248	18	AAW19783	Human multiple reg
42	40	38.1	45	20	AAU12193	Human 5' EST sece
43	40	38.1	137	19	AAU86157	S. pneumoniae deri
44	40	38.1	146	22	AAU1759	Human polypeptide
45	40	38.1	165	21	AAU24840	Plant SDF encoded

ALIGNMENTS

RESULT 1
AAW51092
ID AAW51092 standard; Protein; 286 AA.
XX
AC AAW51092;
XX
DT 14-SEP-1998 (first entry)
XX
DE Ehrlichia chafeensis VSA2 protein.
XX
KW MAP1 homologue; variable surface antigen; VSA2; rickettsia;
KW DNA vaccine.
XX
OS Ehrlichia chafeensis.
XX
FH Key Location/Qualifiers
FT Peptide 1..25
FT /note= "putative signal peptide"
XX
PN WO9816554-A1.
XX
PD 23-APR-1998.
XX
PF 17-OCT-1997; 97WO-US19044.
XX
PR 17-OCT-1996; 96US-0733230.
XX
PA (UYF) UNIV FLORIDA.
XX
PI Barbet AF, Burrig MJ, Ganta RR, McGuire TC;
PI Nylka A, Rurangirwa FR;
XX
DR WPI; 1998-251232/22.
DR N-PSDB; AAV07179.

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NTTGVFGIEQDWDRCVIS 19
 |||||
 Db 60 nttgvfgieqdwdrvcvis 78

RESULT 4

ID AAU04196
 AAU04196 standard; Protein; 286 AA.

XX AC AAU04196;

XX DT 23-OCT-2001 (first entry)

XX DE Variable surface antigen 2 (VSA2) from Ehrlichia chaffeensis.

XX KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia; infection; heartwater; diagnostic; variable surface antigen; VSA.

XX OS Ehrlichia chaffeensis.

XX PN US6251872-B1.

XX XX 26-JUN-2001.

XX PF 17-OCT-1997; 97US-09533326.

XX PR 17-OCT-1996; 96US-0733230.

XX PA (UYFL) UNIV FLORIDA.

XX PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;

XX PI Rurangirwa FR, Mahan SM, Bowie MV, Allenan AR;

XX XX WPI; 2001-424487/45.

DR N-PSDB; AAS07578.

XX PT New MAP2 genes and polypeptides useful as vaccines for conferring immunity to human and animal rickettsial diseases, e.g. heartwater, or as molecular markers in nucleic acid analysis procedures -

XX PS Example 3; Fig 2A-2B; 30pp; English.

XX CC The sequence represents the amino acid sequence of variable surface antigen 2 (VSA2) isolated from Ehrlichia chaffeensis, which has similarity to major antigen protein (MAP). The MAP polynucleotides and polypeptides are useful as vaccines for conferring immunity to rickettsia infection, including Cowdria ruminantium causing heartwater. The MAP polynucleotides may be used as molecular markers in nucleic acid analysis procedures, and to produce the MAP polypeptides, which may be used to raise antibodies that are reactive with the polypeptides. CC The nucleic acids may further be used as probes to identify complementary sequences within other nucleic acid molecules or genomes, where such probes can be applied to identify or distinguish infectious strains of organisms in diagnostic procedures or in rickettsial research where identification of particular organisms or strains is needed.

XX SQ Sequence 286 AA;

Query Match 100.0%; Score 105; DB 22; Length 286;
 Best Local Similarity 100.0%; Pred. No. 1.7e-08;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NTTGVFGIEQDWDRCVIS 19
 |||||
 Db 60 nttgvfgieqdwdrvcvis 78

RESULT 5

AAW51094

ID AAW51094 standard; Protein; 280 AA.

XX AC AAW51094;

XX DT 14-SEP-1998 (first entry)

XX DE Ehrlichia chaffeensis VSA4 protein.

XX KW MAP1 homologue; variable surface antigen; VSA4; rickettsia; DNA vaccine.

XX OS Ehrlichia chaffeensis.

XX PH Key Location/Qualifiers

XX FT Peptide 1..25

XX FT /note= "putative signal peptide"

XX XX WO9816554-A1.

XX PN 23-APR-1998.

XX PF 17-OCT-1997; 97WO-US19044.

XX PR 17-OCT-1996; 96US-0733230.

XX PA (UYFL) UNIV FLORIDA.

XX PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;

XX PI Nyika A, Rurangirwa FR;

XX DR WPI; 1998-251232/22.

XX DR N-PSDB; AAV07179.

XX PT Composition containing nucleic acid encoding rickettsial antigen - useful for, e.g. stimulating protective immune response in humans or animals

XX PS Claim 3; Fig 2B; 39pp; English.

XX CC This is the full-length variable surface antigen VSA4 protein of Ehrlichia chaffeensis. Its amino acid sequence was deduced from a partial open reading frame (ORF4) of a genomic locus (see AAV07179) of E. chaffeensis that was obtained on the basis of homology to the major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium. CC This genomic locus included 5 ORFs encoding similar, but non-identical proteins (see AAW51091-95). A claimed composition comprises a nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99) that elicits a protective immune response against a rickettsial pathogen. The nucleic acid is used, in human or veterinary medicine, in vaccines to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic polypeptides can be used diagnostically to detect antibodies associated with Ehrlichia infection (claimed).

XX SQ Sequence 280 AA;

Query Match 73.3%; Score 77; DB 19; Length 280;
 Best Local Similarity 68.4%; Pred. No. 0.00037;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NTTGVFGIEQDWDRCVIS 19
 |||||
 Db 60 nttgvfgieqdwdrvcvis 78

RESULT 6

AAB36188

ID AAB36188 standard; Protein; 280 AA.

XX AC AAB36188;

XX DT 02-MAR-2001 (first entry)

XX DE Ehrlichia chaffeensis partial VSA4.
 XX KW Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAPI;
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
 KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
 XX 3gдорf3.
 XX OS Ehrlichia chaffeensis.
 XX XX WO200065063-A2.
 XX PN 02-NOV-2000.
 XX PD 21-APR-2000; 2000WO-US10886.
 XX PF 22-APR-1999; 99US-0130725.
 XX PR (UYEL) UNIV FLORIDA.
 XX PA Barbet AF, Bowie MV, Ganta RR, Burrigide MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
 XX WPI; 2000-679675/66.
 XX DR N-PSDB; AAC68705.
 XX XX New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
 PT
 XX Claim 3; Page 45-46; 63pp; English.
 XX The present sequence shows a high degree of similarity to the major
 XX antigenic protein 1 (MAPI) of Ehrlichia sp. The MAPI gene may be
 CC used in a vaccines to protect animals or humans against rickettsial
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid vaccines
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
 CC and 3gдорf3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.
 XX Sequence 280 AA;
 SQ

Query Match 73.3%; Score 77; DB 21; Length 280;
 Best Local Similarity 68.4%; Pred. No. 0.00037;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NTTVGVGIEQDWDRCVIS 19
 |||:||||:|||||
 Db 60 nttigvfgikqdwgdstis 78

RESULT 7
 AAU04198
 ID AAU04198 standard; Protein; 280 AA.
 XX AC AAU04198;
 XX DT 23-OCT-2001 (first entry)
 XX DE Variable surface antigen 4 (VSA4) from Ehrlichia chaffeensis.
 XX KW Major antigenic protein; MAPI; vaccine; immunogenic; rickettsia;
 KW infection; heartwater; diagnostic; variable surface antigen; VSA.
 XX OS Ehrlichia chaffeensis.
 XX PA

PN US6251872-B1.
 XX 26-JUN-2001.
 XX PF 17-OCT-1997; 97US-0953326.
 XX PR 17-OCT-1996; 96US-0733230.
 XX PA (UYEL) UNIV FLORIDA.
 XX XX Barbet AF, Ganta RR, McGuire TC, Burrigide MJ, Nyika A;
 PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
 XX WPI; 2001-424487/45.
 XX DR N-PSDB; AAS07578.
 XX XX New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures
 PT
 XX Example 3; Fig 2A-2B; 30pp; English.
 XX The sequence represents the amino acid sequence of variable surface
 CC antigen 4 (VSA4) isolated from Ehrlichia chaffeensis, which
 CC has similarity to major antigen protein (MAPI). The MAPI polynucleotides
 CC and polypeptides are useful as vaccines for conferring immunity to
 CC rickettsia infection, including Cowdria ruminantium causing heartwater.
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid
 CC analysis procedures, and to produce the MAP polypeptides, which may
 CC be used to raise antibodies that are reactive with the polypeptides.
 CC The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious
 CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is
 CC needed.
 XX Sequence 280 AA;
 SQ

Query Match 73.3%; Score 77; DB 22; Length 280;
 Best Local Similarity 68.4%; Pred. No. 0.00037;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NTTVGVGIEQDWDRCVIS 19
 |||:||||:|||||
 Db 60 nttigvfgikqdwgdstis 78

RESULT 8
 AAY06948
 ID AAY06948 standard; Protein; 280 AA.
 XX AC AAY06948;
 XX DT 05-JUL-1999 (first entry)
 XX DE E. chaffeensis OMP-1F protein.
 XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.
 XX OS Ehrlichia chaffeensis.
 XX PN WO9913720-A1.
 XX PD 25-MAR-1999.
 XX PF 18-SEP-1998; 98WO-US19600.
 XX PR 19-SEP-1997; 97US-0059353.
 XX PA (OHIS) UNIV OHIO STATE.

XX Ohashi N, Rikihisa Y;
PI WPI; 1999-254290/21.
DR N-PSDB; AAX34748.
XX
PT Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX
PS Claim 16; Fig 8B; 55pp; English.
XX
CC The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the p30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 280 AA;

Query Match 70.5%; Score 74; DB 20; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.0011;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVGIEQDWRDRCVIS 19
Db 60 nttgvgfqlkqwdgstis 78

RESULT 9
AAY06942
ID AAY06942 standard; Protein; 256 AA.
XX
AC AAY06942;
XX
DT 05-JUL-1999 (first entry)
XX
DE E. chaffeensis p28 protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
OS Ehrlichia chaffeensis.
XX
PN W09913720-A1.
XX
PD 25-MAR-1999.
XX
PF 18-SEP-1998; 98WO-US19600.
XX
PR 19-SEP-1997; 97US-0059353.
XX
PA (OHIS) UNIV OHIO STATE.
XX
PI Ohashi N, Rikihisa Y;
XX
DR WPI; 1999-254290/21.
DR N-PSDB; AAX34742.
XX
PT Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX
PS Claim 18; Fig 1; 55pp; English.
XX
CC The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the p30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 256 AA;

Query Match 69.5%; Score 73; DB 20; Length 256;
Best Local Similarity 68.4%; Pred. No. 0.0014;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGIEQDWRDRCVIS 19
Db 34 nttgvgfqlkqwdgsais 52

RESULT 10
AAW51095
ID AAW51095 standard; Protein; 276 AA.
XX
AC AAW51095;
XX
DT 14-SEP-1998 (first entry)
XX
DE Ehrlichia chaffeensis VSA5 protein (partial sequence).
XX
KW MAP1 homologue; variable surface antigen; VSA5; rickettsia;
KW DNA vaccine.
XX
OS Ehrlichia chaffeensis.
XX
FH Key Location/Qualifiers
FT Peptide 1..25 /note= "putative signal peptide"
FT
XX
PN W09816554-A1.
XX
PD 23-APR-1998.
XX
PF 17-OCT-1997; 97WO-US19044.
XX
PR 17-OCT-1996; 96US-0733230.
XX
PA (UYFL) UNIV FLORIDA.
XX
PI Barbet AF, Burr ridge MJ, Ganta RR, Mahan SM, McGuire TC;
PI Nyika A, Rurangirwa FR;
XX
DR WPI; 1998-251232/22.
DR N-PSDB; AAV07179.
XX
PT Composition containing nucleic acid encoding rickettsial antigen -
PT useful for, e.g. stimulating protective immune response in humans or
PT animals
XX
PS Claim 3; Fig 2B; 39pp; English.

XX This is the near full-length variable surface antigen VSA5 protein
CC of Ehrlichia chaffeensis; it lacks about 5-7 C-terminal amino acid
CC residues. The VSA5 amino acid sequence was deduced from a partial
CC open reading frame (ORF5) of a genomic locus (see AAV07179) of E.
CC chaffeensis that was obtained on the basis of homology to the major
CC antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium. This
CC genomic locus included 5 ORFs encoding similar, but non-identical
CC proteins (see AAW51091-95). A claimed composition comprises a
CC nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99)
CC that elicits a protective immune response against a rickettsial
CC pathogen. The nucleic acid is used, in human or veterinary
CC medicine, in vaccines to protect against Rickettsia, Ehrlichia,
CC Anaplasma and Cowdria species. The Ehrlichia antigenic
CC polypeptides can be used diagnostically to detect antibodies
CC associated with Ehrlichia infection (claimed).
XX
SQ Sequence 276 AA;

Query Match 69.5%; Score 73; DB 19; Length 276;
Best Local Similarity 68.4%; Pred. No. 0.0015;

Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTVGFGIEQDWDRCVIS 19
|||||:|:|:|
Db 59 nttvgvfgikqndgsais 77

RESULT 11

AA36189
ID AAB36189 standard; Protein; 276 AA.

XX AC AAB36189;

XX XX 02-MAR-2001 (first entry)

XX XX Ehrlichia chaffeensis partial VSA5.

XX KW Ehrlichia chaffeensis: VSA5; variable surface antigen 5; MAP1;
KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
XX 3gdorf3.

XX OS Ehrlichia chaffeensis.

XX PN WC200065063-A2.

XX PD 02-NOV-2000.

XX PF 21-APR-2000; 2000WO-US10886.

XX PR 22-APR-1999; 99US-0130725.

XX PA (UYFL) UNIV FLORIDA.

XX PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
XX N-PSDB; AAC68706.

XX PT New polynucleotides useful as DNA vaccines for conferring immunity to
PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

XX PS Claim 3; Page 47; 63pp; English.

XX CC The present sequence shows a high degree of similarity to the major
CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
CC used in a vaccines to protect animals or humans against rickettsial
CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
CC protective against the rickettsial pathogen. The nucleic acid vaccines
CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
CC polypeptides are useful for detecting antibodies associated with
CC infection by a rickettsial pathogen whilst the polynucleotides may be
CC used to detect the presence of rickettsial nucleic acids.

XX SQ Sequence 276 AA;

Query Match 69.5%; Score 73; DB 21; Length 276;

Best Local Similarity 68.4%; Pred. No. 0.0015;

Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTVGFGIEQDWDRCVIS 19
|||||:|:|:|
Db 59 nttvgvfgikqndgsais 77

RESULT 12

AAU04199
ID AAU04199 standard; Protein; 276 AA.

XX AC AAU04199;

XX DT 23-OCT-2001 (first entry)

XX DE Variable surface antigen 5 (VSA5) from Ehrlichia chaffeensis.

XX KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
KW infection; heartwater; diagnostic; variable surface antigen; VSA.

XX OS Ehrlichia chaffeensis.

XX PN US6251872-B1.

XX PD 26-JUN-2001.

XX PF 17-OCT-1997; 97US-0953326.

XX PR 17-OCT-1996; 96US-0733230.

XX PA (UYFL) UNIV FLORIDA.

XX PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;

XX DR WPI; 2001-424487/45.

XX DR N-PSDB; AAS07578.

XX PT New MAP2 genes and polypeptides useful as vaccines for conferring
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
PT as molecular markers in nucleic acid analysis procedures

XX PS Example 3; Fig 2A-2B; 30pp; English.

XX CC The sequence represents the amino acid sequence of variable surface
CC antigen 5 (VSA5) isolated from Ehrlichia chaffeensis, which
CC has similarity to major antigen protein (MAP). The MAP polynucleotides
CC and polypeptides are useful as vaccines for conferring immunity to
CC rickettsia infection, including Cowdria ruminantium causing heartwater.
CC The MAP polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures, and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish infectious
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
CC needed.

XX SQ Sequence 276 AA;

Query Match 69.5%; Score 73; DB 22; Length 276;

Best Local Similarity 68.4%; Pred. No. 0.0015;

Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTVGFGIEQDWDRCVIS 19
|||||:|:|:|
Db 59 nttvgvfgikqndgsais 77

RESULT 13

AAW51089

ID AAW51089 standard; Protein; 280 AA.

XX AC AAW51089;

XX DT 14-SEP-1998 (first entry)

XX DE Ehrlichia chaffeensis major antigenic protein 1 (MAP1).

KW MAP1 gene; major antigenic protein 1; rickettsia; DNA vaccine.

XX Ehrlichia chaffeensis.

PN WO9816554-A1.

XX 23-APR-1998.

XX 17-OCT-1997; 97WO-US19044.

XX 17-OCT-1996; 96US-0733230.

XX (UYFL) UNIV FLORIDA.

PI Barbet AF, Burrldge MJ, Ganta RR, Mahan SM, McGuire TC;

PI Nyika A, Rurangirwa FR;

DR WPI; 1998-251232/22.

DR N-PSDB; AAV07177.

XX Composition containing nucleic acid encoding rickettsial antigen -
PT useful for, e.g. stimulating protective immune response in humans or
PT animals

XX Claim 3; Page 18-19; 39pp; English.

CC This polypeptide comprises the major antigen protein 1 gene (MAP1)
of Ehrlichia chaffeensis. It is encoded by the MAP1 gene (see
CC AAV07177). A claimed composition comprises a nucleic acid (see
CC AAV07176-82) encoding a polypeptide (see AAV51088-99) that elicits a
CC protective immune response against a rickettsial pathogen. The
CC nucleic acid is used, in human or veterinary medicine, in vaccines
CC to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria
CC species. The nucleic acid does not replicate in the host but
CC remains episomal and capable of expressing polypeptide for at least
CC 19 mth. The Ehrlichia antigenic polypeptides can be used
CC diagnostically to detect antibodies associated with Ehrlichia
CC infection (claimed).

XX Sequence 280 AA;

Query Match 69.5%; Score 73; DB 19; Length 280;

Best Local Similarity 68.4%; Pred. No. 0.0015;

Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NTTGVFGIEQDWDRCVIS 19

Db 60 nttgvfglkqndgsais 78

RESULT 14

AAB36183

ID AAB36183 standard; Protein; 280 AA.

XX AAB36183;

XX 02-MAR-2001 (first entry)

XX Ehrlichia chaffeensis MAP1.

DE Ehrlichia chaffeensis; MAP1; major antigenic protein 1; antirickettsial;
KW vaccine; gene therapy; Rickettsia; Cowdria; Anaplasma; map2; lhworf3;
KW 4hworf1; 18hworf1; 3gdoorf3.

XX Ehrlichia chaffeensis.

XX WO200065063-A2.

XX 02-NOV-2000.

XX 21-APR-2000; 2000WO-US10886.

XX

PR

XX

PA

XX

XX

PI

PI

XX

XX

DR

DR

XX

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

22-APR-1999; 99US-0130725.

(UYFL) UNIV FLORIDA.

Barbet AF, Bowie MV, Ganta RR, Burrldge MJ, Mahan SM, McGuire TC;
Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Allemen AR;

WPI; 2000-679675/66.

N-PSDB; AAC68700.

New polynucleotides useful as DNA vaccines for conferring immunity to
rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

Claim 3; Page 35-36; 63pp; English.

The present sequence is given in a specification relating to nucleic
acid vaccines containing genes to protect animals or humans against
rickettsial diseases caused by a organisms of Rickettsia sp., Ehrlichia
sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
protective against the rickettsial pathogen. The vaccine comprises the
major antigenic protein 1 (MAP1) gene or the major antigenic protein 2
(MAP2) gene of rickettsial pathogens. The nucleic acid vaccines can be
driven by the human cytomegalovirus (HCMV) enhancer-promoter. Cowdria
ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1 and
3gdoorf3 may be used in therapeutic and diagnostic applications. The
polypeptides are useful for detecting antibodies associated with
infection by a rickettsial pathogen whilst the polynucleotides may be
used to detect the presence of rickettsial nucleic acids.

Sequence 280 AA;

Query Match 69.5%; Score 73; DB 21; Length 280;

Best Local Similarity 68.4%; Pred. No. 0.0015;

Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NTTGVFGIEQDWDRCVIS 19

Db 60 nttgvfglkqndgsais 78

RESULT 15

AAU04193

ID AAU04193 standard; Protein; 280 AA.

XX AAU04193;

XX 23-OCT-2001 (first entry)

XX Major antigenic protein 1 (MAP1) from Ehrlichia chaffeensis.

XX Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
infection; heartwater; diagnostic.

XX Ehrlichia chaffeensis.

XX US6251872-B1.

XX 26-JUN-2001.

XX 17-OCT-1997; 97US-0953326.

XX 17-OCT-1996; 96US-0733230.

XX (UYFL) UNIV FLORIDA.

XX Barbet AF, Ganta RR, McGuire TC, Burrldge MJ, Nyika A;

XX Rurangirwa FR, Mahan SM, Bowie MV, Allemen AR;

XX WPI; 2001-424487/45.

XX N-PSDB; AAS07576.

DR

XX New MAP2 genes and polypeptides useful as vaccines for conferring
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
PT as molecular markers in nucleic acid analysis procedures -
XX
XX Disclosure; Column 15-17: 30pp; English.
XX
XX The sequence represents the amino acid sequence of major antigenic
CC protein 1 (MAP1) from Ehrlichia chaffeensis. The MAP polynucleotides and
CC polypeptides are useful as vaccines for conferring immunity to rickettsia
CC infection, including Cowdria ruminantium causing heartwater. The MAP
CC polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish infectious
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
CC needed.
XX
XX Sequence 280 AA;

Query Match 69.5%; Score 73; DB 22; Length 280;
Best Local Similarity 68.4%; Pred. NO. 0.0015;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 NTTGVFGIEQDWDRCVIS 19
|||||||:|:|
Db 60 ntgvfglkpwwdsais 78

Search completed: March 14, 2002, 09:09:38
Job time: 429 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:24:10 ; Search time 103.37 Seconds
(without alignments)
25.471 Million cell updates/sec

Title: US-09-765-739A-6

Perfect score: 97

Sequence: 1 NPTVALYGLKQDWEGISS 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL17:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phase:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	97	100.0	278	2	O52106 ehrlichia c
2	88	90.7	280	2	O52104 ehrlichia c
3	62	63.9	280	2	O85357 ehrlichia c
4	59	60.8	246	2	O9RH35 ehrlichia c
5	59	60.8	276	2	O85817 ehrlichia c
6	59	60.8	280	2	O52107 ehrlichia c
7	59	60.8	288	2	O92G32 ehrlichia c
8	58	59.8	276	2	O85358 ehrlichia c
9	58	59.8	281	2	O9AC19 ehrlichia c
10	56	57.7	272	2	O9AMP6 ehrlichia sp.
11	56	57.7	280	2	O9F473 ehrlichia c
12	56	57.7	280	2	O9ADV3 ehrlichia c
13	56	57.7	284	2	O46327 ehrlichia c
14	56	57.7	284	2	O9AFA1 ehrlichia c
15	54	55.7	280	2	O85816 ehrlichia c
16	54	55.7	280	2	O92GM9 ehrlichia c
17	53	54.6	278	2	O9R8A9 ehrlichia c
18	53	54.6	278	2	O9R8A8 ehrlichia c
19	53	54.6	278	2	O9R8A7 ehrlichia c

20	53	54.6	278	2	O9R8A6 ehrlichia c
21	53	54.6	278	2	O9R8A5 ehrlichia c
22	53	54.6	278	2	O9R3J3 ehrlichia c
23	53	54.6	278	2	O9F472 ehrlichia c
24	53	54.6	307	2	O9ZGJ1 ehrlichia c
25	52	53.6	265	2	O9AF99 ehrlichia c
26	52	53.6	281	2	O46328 ehrlichia c
27	52	53.6	281	2	O9S6H1 ehrlichia c
28	52	53.6	281	2	O9S6H0 ehrlichia c
29	52	53.6	287	2	O46329 ehrlichia c
30	52	53.6	287	2	O46331 ehrlichia c
31	52	53.6	287	2	O9R425 ehrlichia c
32	52	53.6	290	2	O46330 ehrlichia c
33	52	53.6	290	2	O46333 ehrlichia c
34	51	52.6	270	2	O9AF98 ehrlichia c
35	51	52.6	286	2	O52105 ehrlichia c
36	51	52.6	290	2	O46324 ehrlichia c
37	51	52.6	290	2	O46332 ehrlichia c
38	51	52.6	290	2	O9AEU3 ehrlichia c
39	48	49.5	133	2	O85360 ehrlichia c
40	48	49.5	271	2	O9AFA0 ehrlichia c
41	48	49.5	276	2	O9F475 ehrlichia c
42	48	49.5	283	2	O9F474 ehrlichia c
43	47	48.5	308	5	O9W3R8 ehrlichia c
44	47	48.5	904	10	O9FU09 ehrlichia c
45	45.5	46.9	122	2	O25081 ehrlichia c

ALIGNMENTS

RESULT 1	
O52106	PRELIMINARY; PRT; 278 AA.
ID O52106	
AC O52106;	
DT 01-JUN-1998 (TREMBlrel. 06, Created)	
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)	
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)	
DE OMP-1E.	
GN OMP-1E.	
OS Ehrlichia chaffeensis.	
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;	
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.	
OX NCBI TaxID=945;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=ARKANSAS;	
RA MEDLINE=98084465; PubMed=9423849;	
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;	
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis	
RT are encoded by a polymorphic multigene family.";	
RL Infect. Immun. 66:132-139(1998).	
RN [2]	
RP SEQUENCE FROM N.A.	
RC STRAIN=ARKANSAS;	
RX MEDLINE=98321180; PubMed=9647746;	
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,	
RA Alleman A.R.;	
RT "Molecular characterization of a 28 kDa surface antigen gene family of	
RT the tribe Ehrlichieae.";	
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).	
DR EMBL; U72291; AAC02939.1;	
DR EMBL; AF062761; AAC26719.1;	
DR InterPro; IPR002566; Surface_Ag_msp4.	
DR Pfam; PF01611; Surface_Ag_2; 1.	
SQ SEQUENCE 278 AA; 30543 MW; E321E3CA259B87FD CRC64;	

Query Match 100.0%; Score 97; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 7.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 NPTVALYGLKQDWEGISS 18

Db 60 NPTVALYGLKQDWEGISS 77

RESULT 2
ID O52104 PRELIMINARY; PRT; 280 AA.
AC O52104;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE OMP-1C (28 KDA MAJOR SURFACE ANTIGEN-1).
GN OMP-1C.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Chashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE OF 238-280 FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Allenan A.R.;
RT "Molecular characterization of a 28 kda surface antigen gene family of
the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL; U72291; AAC02937.1; -;
DR EMBL; AF062761; AAC26717.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30322 MW; BD835E491086DF01 CRC64;

Query Match 90.7%; Score 88; DB 2; Length 280;
Best Local Similarity 83.3%; Pred. No. 2.3e-06;
Matches 15; Conservative .2; Mismatches 1; Indels 0; Gaps 0;

OY 1 NPTVALYGLKQDWEGISS 18
Db 60 NPTVALYGLKQDWNGVSA 77

RESULT 3
ID O85357 PRELIMINARY; PRT; 280 AA.
AC O85357;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 28 KDA MAJOR SURFACE ANTIGEN-4.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Allenan A.R.;
RT "Molecular characterization of a 28 kda surface antigen gene family of
the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL; AF062761; AAC26720.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.

DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30743 MW; FBB841DAF08E4DC CRC64;

Query Match 63.9%; Score 62; DB 2; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.038;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 NPTVALYGLKQDWEG 15
Db 60 NTTIGVFLKQDWG 74

RESULT 4
ID Q9RH35 PRELIMINARY; PRT; 246 AA.
AC Q9RH35;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR (FRAGMENT).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ST.VINCENT;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
DR EMBL; AF077735; AAC31548.1; -;
DR InterPro; IPR001702; Gram_neg_porin.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
DR PRINTS; PR00182; ECOLNEIPORIN.
FT NON_TER 1
SQ SEQUENCE 246 AA; 26884 MW; C9776392C5129A2F CRC64;

Query Match 60.8%; Score 59; DB 2; Length 246;
Best Local Similarity 60.0%; Pred. No. 0.1;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 NPTVALYGLKQDWEG 15
Db 29 NTAGVFLKQDWG 43

RESULT 5
ID O85817 PRELIMINARY; PRT; 276 AA.
AC O85817;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAPULPA;
RA Yu X.-J., Walker D.H.;
RT "Ehrlichia chaffeensis 28 kda outer membrane protein.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF077734; AAC31547.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30027 MW; 2FD3698FCF1F60BE CRC64;

```

Query Match          60.8%; Score 59; DB 2; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.12;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
DB 59 NTTAGVFGGLKQDWGD 73

RESULT 6
ID O52107 PRELIMINARY; PRT; 280 AA.
AC O52107;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE OMP-1F.
GN OMP-1F.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RL Infect. Immun. 66:132-139(1998).
DR EMBL; U72291; AAC02940.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30731 MW; CCAAG6C34E2AF393E CRC64;

Query Match          60.8%; Score 59; DB 2; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.12;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
DB 60 NTTGVFGGLKQDWGD 74

RESULT 7
ID Q9ZGJ2 PRELIMINARY; PRT; 288 AA.
AC Q9ZGJ2;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 30-KDA MAJOR OUTER MEMBRANE PROTEIN (P28-8).
GN P30 OR P28-8.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JAKE;

Query Match          60.8%; Score 59; DB 2; Length 288;
Best Local Similarity 60.0%; Pred. No. 0.12;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
DB 60 NTTGVFGGLKQDWGD 74

RESULT 8
ID O85358 PRELIMINARY; PRT; 276 AA.
AC O85358;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 28 KDA MAJOR SURFACE ANTIGEN-5 (FRAGMENT).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichieae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL; AF062761; AAC26716.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
DR NON_TER 276 276
SQ SEQUENCE 276 AA; 29782 MW; A9AB2A92263CA4EA CRC64;

Query Match          59.8%; Score 58; DB 2; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.17;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
DB 59 NTTGVFGGLKQDWGD 73

RESULT 9
ID Q9ACI9 PRELIMINARY; PRT; 281 AA.
AC Q9ACI9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

```

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE MAJOR OUTER MEMBRANE PROTEIN P28.
 GN P28.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ARKANSAS;
 RX MEDLINE=98084465; PubMed=9423849;
 RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
 RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
 RL are encoded by a polymorphic multigene family.";
 RN Infect. Immun. 66:132-139(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ARKANSAS;
 RX MEDLINE=21153566; PubMed=11254561;
 RA Ohashi N., Rikihisa Y., Unver A.;
 RT "Analysis of transcriptionally active Gene Clusters of Major Outer
 RN Membrane Protein Multigene Family in Ehrlichia canis and E.
 RT chaffeensis.";
 RL Infect. Immun. 69:2083-2091(2001).
 RN EMBL: U22911;
 SQ SEQUENCE 261 AA; 30343 MW; A99E5F7C4459AA9A CRC64;

Query Match 59.8%; Score 58; DB 2; Length 281;
 Best Local Similarity 60.0%; Pred. No. 0.17;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
 Db 59 NTTVGVFGLKQNDG 73

RESULT 10
 Q9AMF6 PRELIMINARY; PRT; 272 AA.
 ID Q9AMF6
 AC Q9AMF6;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE MAJOR ANTIGENIC PROTEIN MAP1 (FRAGMENT).
 GN MAP1.
 OS Cowdria sp. 'South African canine'.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichiae; Cowdria.
 OX NCBI_TaxID=152574;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Allisopp M.T., Allisopp B.A.;
 RT "A novel Ehrlichia detected in dogs in South Africa.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF325176; AAK14320.1;
 FT NON_TER 1
 FT NON_TER 272 272
 SQ SEQUENCE 272 AA; 29225 MW; 43CEB95DEDD55BDE CRC64;

Query Match 57.7%; Score 56; DB 2; Length 272;
 Best Local Similarity 50.0%; Pred. No. 0.35;
 Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEGISS 18
 Db 57 TRAVEGLKQDWGVKT 72

RESULT 11
 Q9F473 PRELIMINARY; PRT; 280 AA.
 ID Q9F473

AC Q9F473;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 GN P28-6.
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAKE;
 RX MEDLINE=99242757; PubMed=10225842;
 RA McBride J.W., Yu X.J., Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive
 RN 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 RT antigen.";
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAKE;
 RX MEDLINE=20432107; PubMed=10974556;
 RA McBride J.W., Yu X.J., Walker D.H.;
 RT "A conserved, transcriptionally active p28 multigene locus of
 RN Ehrlichia canis.";
 RL Gene 254:245-252(2000).
 DR EMBL: AF082744; AAG14361.1;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 280 AA; 30762 MW; BE284A4B94FE3123 CRC64;

Query Match 57.7%; Score 56; DB 2; Length 280;
 Best Local Similarity 60.0%; Pred. No. 0.36;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
 Db 59 NSTGVFGLKHDWG 73

RESULT 12
 Q9ADV3 PRELIMINARY; PRT; 280 AA.
 ID Q9ADV3
 AC Q9ADV3;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE MAJOR OUTER MEMBRANE PROTEIN P30-2.
 GN P30-2.
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Ehrlichiae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OKLAHOMA;
 RX MEDLINE=98371112; PubMed=9705412;
 RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
 RT "Cloning and characterization of multigenes encoding the
 RN immunodominant 30-kilodalton major outer membrane proteins of
 RT Ehrlichia canis and application of the recombinant protein for
 RT serodiagnosis.";
 RL J. Clin. Microbiol. 36:2671-2680(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OKLAHOMA;
 RX MEDLINE=21153566; PubMed=11254561;
 RA Ohashi N., Rikihisa Y., Unver A.;
 RT "Analysis of transcriptionally active Gene Clusters of Major Outer
 RN Membrane Protein Multigene Family in Ehrlichia canis and E.
 RT chaffeensis.";

RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF078553; AAK28699.1; -.
SQ SEQUENCE 280 AA; 30803 MW; 27238BE1C7E68A91 CRC64;

Query Match 57.7%; Score 56; DB 2; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.36;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
| | | | | | | | | |
Db 59 NSTVGVFGLKHDWNG 73

RESULT 13
ID Q46327 PRELIMINARY; PRT; 284 AA.
AC Q46327;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE MAJOR ANTIGENIC PROTEIN.
GN MAP1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Cowdria.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SENEGAL STOCK;
RX MEDLINE=94178956; PubMed=8132352;
RA van Vliet A.H., Jongejan F., van Kleef M., Van der Zeijst B.A.;
RT "Molecular cloning, sequence analysis, and expression of the gene
RT encoding the immunodominant 32-kilodalton protein of Cowdria
RT ruminantium.";
RL Infect. Immun. 62:1451-1456(1994).
DR EMBL; X74250; CAA52309.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 284 AA; 30634 MW; 53228A889D28BEB8 CRC64;

Query Match 57.7%; Score 56; DB 2; Length 284;
Best Local Similarity 50.0%; Pred. No. 0.37;
Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEGISS 18
| | | | | | | | | |
Db 63 TRAVFGLKKDWDGVT 78

RESULT 14
ID Q9AFAL PRELIMINARY; PRT; 284 AA.
AC Q9AFAL;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE MAJOR ANTIGENIC PROTEIN 1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Cowdria.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALL-3;
RA Bensaid A., Maillard J.-C., Chantal I., Perez J.-M., Martinez D.;
RT "Cowdria ruminantium major antigenic protein 1 (map1) gene variants
RT are not geographically constrained."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF355200; AAK27216.1; -.
SQ SEQUENCE 284 AA; 30720 MW; B0D3AEB9F9AB09C1 CRC64;

Query Match 57.7%; Score 56; DB 2; Length 284;
Best Local Similarity 50.0%; Pred. No. 0.37;
Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEGISS 18
| | | | | | | | | |
Db 63 TRAVFGLKKDWDGVT 78

RESULT 15
ID O85816 PRELIMINARY; PRT; 280 AA.
AC O85816;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE OUTER MEMBRANE PROTEIN P28 PRECURSOR.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Ehrlichieae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=91HE17;
RA Yu X.-J., Walker D.H.;
RT "Ehrlichia chaffeensis 28 kDa outer membrane protein.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077732; AAC31545.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30277 MW; 91C54AC78507A63F CRC64;

Query Match 55.7%; Score 54; DB 2; Length 280;
Best Local Similarity 61.5%; Pred. No. 0.76;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEG 15
| | | | | | | | | |
Db 61 TAGVFLKQDWG 73

Search completed: March 14, 2002, 09:24:10
Job time: 981 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2002, 09:24:52 ; Search time 32.82 seconds
(without alignments)
20.109 Million cell updates/sec

Title: US-09-765-739A-6
Perfect score: 97
Sequence: 1 NPTVALYGLKQDWEGISS 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	42	43.3	1023	1 TSCC_PSEAM	P55019 pseudopleur
2	41	42.3	169	1 FLAV_ANASP	P11241 anabaena sp
3	41	42.3	452	1 YK97_MYCTU	Q10706 mycobacteri
4	41	42.3	544	1 PYRG_AZOB	P28595 azospirillu
5	41	42.3	918	1 OA1S_NEUCR	P11637 neurospora
6	41	42.3	991	1 DRPL_SCHPO	P40848 schizosacch
7	40	41.2	169	1 FLAV_SYN7	P10340 synecococc
8	40	41.2	239	1 DCOP_BACSU	P25971 bacillus su
9	40	41.2	341	1 Y33B_MYCPN	P75302 mycoplasma
10	40	41.2	661	1 YN84_YEAST	P40345 saccharomyc
11	40	41.2	734	1 PURL_YMMO	O9reg6 zymomonas m
12	40	41.2	1060	1 NKCL_MANSE	Q25479 manduca sex
13	40	41.2	1220	1 DX8_HUMAN	Q14562 homo sapien
14	39.5	40.7	354	1 CARA_METUA	Q58425 methanococc
15	39	40.2	196	1 NODA_RHIS3	P72329 rhizobium s
16	39	40.2	226	1 NODA_AZOC	Q07739 azorhizobi
17	39	40.2	264	1 KRA3_ENTFA	P00554 enterococu
18	39	40.2	283	1 HXA9_FUGRU	Q42506 fugu rubrip
19	39	40.2	286	1 RIPL_MOMCH	P16094 momordica c
20	39	40.2	415	1 PGT_A_SALTY	P06184 salmonella
21	39	40.2	493	1 VG10_BPMD2	O64203 mycobacteri
22	39	40.2	589	1 VP40_SCMVC	P16046 simian cyto
23	39	40.2	643	1 VP40_HSV2	P52369 equine herp
24	39	40.2	688	1 DNLJ_THEMA	O9wxv5 thermotoga
25	39	40.2	732	1 EF2_PYRAB	Q9W128 pyrococcus
26	39	40.2	732	1 EF2_PYRHO	O59521 pyrococcus
27	39	40.2	868	1 PRIT_PORGI	P43158 porphyromon
28	39	40.2	1102	1 TRAA_RHISN	P55418 rhizobium s
29	39	40.2	1191	1 NKCL_SQUAC	P55013 squalus aca
30	38.5	39.7	710	1 CDGT_THERTU	P26827 thermopane
31	38	39.2	184	1 VF50_BPAPS	Q9tip8 bacterioph
32	38	39.2	290	1 AROE_SYNY3	P74591 synecocyst
33	38	39.2	370	1 CYP4_BOVIN	P26882 bos taurus

RESULT 1

TSCC_PSEAM

ID	TSCC_PSEAM	STANDARD;	PRT; 1023 AA.
AC	P55019;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	01-NOV-1997 (Rel. 35, Last annotation update)		
DE	THIAZIDE-SENSITIVE SODIUM-CHLORIDE COTRANSPORTER (NA-CL SYMPORTER).		
GN	SLC12A3 OR TSC.		
OS	Pseudopleuronectes americanus (winter flounder).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;		
OC	Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.		
OX	NCBI_TaxID=8265;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Bladder urothelium;		
RX	MEDLINE=93219361; PubMed=8464884;		
RA	Gamba G., Saltzberg S.N., Lombardi M., Miyanoshita A., Lytton J.,		
RA	Hediger M.A., Brenner B.M., Hebert S.C.;		
RT	"Primary structure and functional expression of a cDNA encoding the		
RT	thiazide-sensitive, electroneutral sodium-chloride cotransporter.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 90:2749-2753(1993).		
CC	FUNCTION: ELECTRICALLY SILENT TRANSPORTER SYSTEM WHICH IS A		
CC	MEDIATOR OF SODIUM AND CHLORIDE REABSORPTION.		
CC	TISSUE SPECIFICITY: INTEGRAL MEMBRANE PROTEIN.		
CC	TISSUE SPECIFICITY: URINARY BLADDER.		
CC	SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	EMBL; L11615; AAA49272.1;		
DR	InterPro; IPR002293; AA_rel_permease.1.		
DR	InterPro; IPR002027; Amino_acid_permease.		
DR	InterPro; IPR002948; NaCl_trnsptr.		
DR	Pfam; PF00324; aa_permeases; 1.		
DR	PRINTS; PR01230; NACLTRNSPORT.		
KW	Transport; Transmembrane; Glycoprotein.		
FT	DOMAIN 1 132		
FT	TRANSMEM 133 153		
FT	POTENTIAL.		
FT	DOMAIN 134 163		
FT	EXTRACELLULAR (POTENTIAL).		
FT	TRANSMEM 164 184		
FT	POTENTIAL.		
FT	DOMAIN 185 215		
FT	CYTOPLASMIC (POTENTIAL).		
FT	TRANSMEM 216 236		
FT	POTENTIAL.		
FT	DOMAIN 237 258		
FT	EXTRACELLULAR (POTENTIAL).		
FT	TRANSMEM 259 279		
FT	POTENTIAL.		
FT	DOMAIN 280 283		
FT	CYTOPLASMIC (POTENTIAL).		
FT	TRANSMEM 284 304		
FT	POTENTIAL.		
FT	DOMAIN 305 336		
FT	EXTRACELLULAR (POTENTIAL).		

Q08752 homo sapien
Q282752 canis famil
Q28377 equus cabal
Q16822 homo sapien
Q60568 homo sapien
P07589 bos taurus
P11276 mus musculu
P03172 herpes simp
Q44776 haemophilus
O9y264 homo sapien
P27036 bacillus oh
P36000 saccharomyc

```

FT TRANSMEM 337 357 POTENTIAL.
FT DOMAIN 358 374 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 375 395 POTENTIAL.
FT DOMAIN 396 451 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 452 472 POTENTIAL.
FT DOMAIN 473 510 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 511 531 POTENTIAL.
FT DOMAIN 532 576 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 577 597 POTENTIAL.
FT DOMAIN 598 745 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 746 766 POTENTIAL.
FT DOMAIN 767 830 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 831 851 POTENTIAL.
FT DOMAIN 852 1023 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 117 82 POLY-ARG.
FT DOMAIN 117 122 POLY-GLU.
FT TRANSMEM 407 412 POLY-SER.
FT DOMAIN 933 937 POLY-GLN.
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 543 543 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1023 AA; 112340 MW; 69AE2D53B8F84D89 CRC64;

Query Match 43.3%; Score 42; DB 1; Length 1023;
Best Local Similarity 58.3%; Pred. No. 39;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 PTVALYGLKODW 13
| | | | |
DB 727 PNVLLMGFKKDW 738

RESULT 2
FLAV_ANASP
ID FLAV_ANASP STANDARD; PRT: 169 AA.
AC P11241;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE FLAVODOXIN.
GN ISIB.
OS Anabaena sp. (strain PCC 7120), and
OS Anabaena sp. (strain PCC 7119).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690; 1168;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC 7120;
RX MEDLINE=89296496; PubMed=2500643;
RA Leonhardt K.G., Straus N.A.;
RT "Sequence of the flavodoxin gene from Anabaena variabilis 7120.";
RL Nucleic Acids Res. 17:4384-4384(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC 7119;
RX MEDLINE=92074973; PubMed=1720613;
RA Fillat M.F., Borrias W.E., Weisbeek P.J.;
RT "Isolation and overexpression in Escherichia coli of the flavodoxin
gene from Anabaena PCC 7119.";
RL Biochem. J. 280:187-191(1991).
RN [3]
RP SEQUENCE OF 1-36.
RC STRAIN-PCC 7119;
RX MEDLINE=90381288; PubMed=2119231;
RA Fillat M.F., Edmondson D.E., Gomez-Moreno C.;
RT "Structural and chemical properties of a flavodoxin from Anabaena PCC
7119.";
RL Biochim. Biophys. Acta 1040:301-307(1990).
RN [4]
RP STRUCTURE BY NMR.
RC STRAIN-PCC 7120;

```

```

RX MEDLINE=91104858; PubMed=2125478;
RA Stockman B.J., Krezel A.M., Markley J.L., Leonhardt K.G.,
RA Straus N.A.;
RT "Hydrogen-1, carbon-13, and nitrogen-15 NMR spectroscopy of Anabaena
7120 flavodoxin: assignment of beta-sheet and flavin binding site
resonances and analysis of protein-flavin interactions.";
RL Biochemistry 29:9600-9609(1990).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RC STRAIN-PCC 7120;
RX MEDLINE=93271891; PubMed=1303762;
RA Rao S.T., Shaffie F., Yu C., Satyshur K.A., Stockman B.J.,
RA Markley J.L., Sundaralingam M.;
RT "Structure of the oxidized long-chain flavodoxin from Anabaena 7120
at 2-A resolution.";
RL Protein Sci. 1:1413-1427(1992).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS).
RC STRAIN-PCC 7120;
RA Burkhardt B.M., Ramakrishnan B., Yan H., Reedstrom R.J., Markley J.L.,
RA Straus N.A., Sundaralingam M.;
RT "Structure of the trigonal form of recombinant oxidized flavodoxin
from Anabaena 7120 at 1.40-A resolution.";
RL Acta Crystallogr. D 51:318-330(1995).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RC STRAIN-PCC 7119;
RX MEDLINE=99318886; PubMed=10388575;
RA Fernandez-Recio J., Romero A., Sancho J.;
RT "Energetics of a hydrogen bond (charged and neutral) and of a
cation-pi interaction in apoflavodoxin.";
RL J. Mol. Biol. 290:319-330(1999).
CC -!- FUNCTION: LOW-POTENTIAL ELECTRON DONOR TO A NUMBER OF REDOX
CC ENZYMES.
CC -!- COFACTOR: FMN.
CC -!- SIMILARITY: BELONGS TO THE FLAVODOXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: S68006; AAB20462.1;
CC PIR: S04600; S04600.
CC PIR: S18374; S18374.
CC PIR: S20298; S20298.
CC PDB: 1RLV; 31-OCT-93.
CC PDB: 1FTG; 23-DEC-96.
CC PDB: 1RCF; 26-JAN-95.
CC PDB: 1QHE; 20-MAY-99.
CC PDB: 1DX9; 10-APR-00.
CC InterPro; IPR001226; Flavodoxin.
CC Pfam; PF00258; flavodoxin; 1.
CC PROSITE; P500201; FLAVODOXIN; 1.
KW Electron transport; Flavoprotein; FMN; 3D-structure.
FT INIT_MET 0 0
FT STRAND 4 8
FT HELIX 14 26
FT TURN 31 35
FT TURN 36 37
FT TURN 41 46
FT STRAND 49 53
FT STRAND 56 57
FT TURN 58 60
FT STRAND 61 62
FT HELIX 64 70
FT TURN 71 72
FT HELIX 73 75
FT TURN 79 80

```

FT STRAND 82 88
 FT TURN 91 96
 FT TURN 98 99
 FT TURN 100 111
 FT TURN 112 113
 FT TURN 115 116
 FT STRAND 120 120
 FT TURN 123 124
 FT TURN 131 132
 FT TURN 133 134
 FT TURN 135 136
 FT STRAND 137 138
 FT STRAND 141 143
 FT TURN 145 147
 FT TURN 149 165
 FT TURN 166 168
 SQ SEQUENCE 169 AA; 18832 MW; BB0BBD0B6DEF9A58 CRC64;

Query Match 42.3%; Score 41; DB 1; Length 169;
 Best Local Similarity 47.1%; Pred. No. 8.7;
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 PTVALYGLKQDWEGISS 18
 II : I: ||||: I
 Db 55 PTWNIGELQSDWEGLYS 71

RESULT 3
 YK97_MYCTU
 ID YK97_MYCTU STANDARD; PRT; 452 AA.
 AC Q10706;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 51.4 KDA PROTEIN RV2097C.
 GN RV2097C OR MT2158 OR MTCY49.37C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: STRONG, TO M.LEPRAE ML1328.
 CC -1- SIMILARITY: TO RHODOCOCUS STRAIN N186/21 HYPOTHETICAL PROTEIN IN
 CC THCR 5'REGION (ORF6) (AC P43484).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z73966; CAA98209.1; -.
 DR EMBL; AE007065; AAK46439.1; -.
 DR TIGR; MT2158; -.
 DR TuberculList; RV2097c; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 452 AA; 51384 MW; 12AF8B7872D2F5EA CRC64;

Query Match 42.3%; Score 41; DB 1; Length 452;
 Best Local Similarity 47.6%; Pred. No. 25;
 Matches 10; Conservative 1; Mismatches 6; Indels 4; Gaps 1;

QY 2 PTVALYGLKQD---WEGISS 18
 I I I I I |||: ||
 Db 161 PKAATYCLSORAEHWEGVSS 181

RESULT 4
 PYRG_AZOBR
 ID PYRG_AZOBR STANDARD; PRT; 544 AA.
 AC P28595;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE CTP SYNTHASE (EC 6.3.4.2) (UTP--AMMONIA LIGASE) (CTP SYNTHETASE).
 GN PYRG.
 OS Azospirillum brasilense.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
 OC Azospirillum.
 OX NCBI_TaxID=192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29145 / SP7;
 RX MEDLINE=94186025; PubMed=8138139;
 RA Zimmer W., Hundsdoergen B.;
 RT "Identification and sequencing of pyrG, the CTP synthetase gene of
 RT Azospirillum brasilense SP7."
 RL FEMS Microbiol. Lett. 115:273-278 (1994).
 CC -1- CATALYTIC ACTIVITY: ATP + UTP + GLUTAMINE = ADP + ORTHOPHOSPHATE +
 CC CTP (AMMONIA CAN REPLACE GLUTAMINE).
 CC -1- ENZYME REGULATION: ACTIVATED BY GTP AND INHIBITED BY CTP.
 CC -1- PATHWAY: LAST STEP IN PYRIMIDINE BIOSYNTHESIS.
 CC -1- SIMILARITY: TO OTHER SPECIES CTP SYNTHASE.
 CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-1 GLUTAMINE
 CC AMIDOTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X67216; CAA47656.1; -.
 DR PIR; S25101; S25101.
 DR InterPro; IPR000991; GATase_1.
 DR Pfam; PF00117; GATase; 1.
 DR PROSITE; PS00442; GATASE_TYPE_I; 1.
 KW Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase.
 FT ACT_SITE 380 380 GATASE (BY SIMILARITY).
 FT ACT_SITE 516 516 GATASE (BY SIMILARITY).
 FT ACT_SITE 518 518 GATASE (BY SIMILARITY).
 SQ SEQUENCE 544 AA; 59908 MW; 13AA1A8B8EB6C5DB CRC64;

Query Match 42.3%; Score 41; DB 1; Length 544;

```
Best Local Similarity 53.8%; Pred. No. 30;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDW 13
   |||||
DB 410 NPVVGLGLMTEW 422

RESULT 5
OALS_NEUCR
ID OALS_NEUCR STANDARD; PRT; 918 AA.
AC P11637;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE QUINATE REPRESSOR.
GN QA-1S.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=89293848; PubMed=2525625;
RA Geever R.F., Huiet L., Baum J.A., Tyler B.M., Patel V.B.,
RT Rutledge B.J., Case M.E., Giles N.H.;
RT "DNA sequence, organization and regulation of the qa gene cluster of
RT Neurospora crassa.";
RL J. Mol. Biol. 207:15-34(1989).
CC -!- FUNCTION: REPRESSOR FOR ENZYMES AND PROTEINS OF QUINATE
CC METABOLISM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14603; CAA32753.1;
DR PIR; S04235; S04255.
DR PIR; E31277; E31277.
DR InterPro; IPR001361; DQKinasase_I.
DR InterPro; IPR000623; Shik_Kinase.
DR InterPro; IPR002907; Shikimate_DH.
DR Pfam; PF01487; DQKinasase_I; 1.
DR Pfam; PF01488; Shikimate_DH; 1.
DR Pfam; PF01202; SKI; 1.
DR Quinate metabolism; Transcription regulation; Repressor; DNA-binding.
SQ SEQUENCE 918 AA; 100580 MW; 67EDA399CBF098B2 CRC64;

Query Match 42.3%; Score 41; DB 1; Length 918;
Best Local Similarity 66.7%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 ALYGLKQDWEGI 16
   |||||
DB 684 ALYGTNDWIGI 695

RESULT 6
DHPI_SCHPO
ID DHPI_SCHPO STANDARD; PRT; 991 AA.
AC P40848;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN DHPI.
GN DHPI OR SPAC26A3.12.
OS Schizosaccharomyces pombe (Fission yeast).

Best Local Similarity 53.8%; Pred. No. 30;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-29.
RC STRAIN=975;
RX MEDLINE=94247347; PubMed=8190062;
RA Sugano S, Shobuike T, Takeda T., Sugino A., Ikeda H.;
RT "Molecular analysis of the dhpi+ gene of Schizosaccharomyces pombe:
RT an essential gene that has a homology to the DST2 and RAI1 genes of
RT Saccharomyces cerevisiae."
RL Mol. Gen. Genet. 243:1-8(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA McLean J., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ESSENTIAL FOR VEGETATIVE CELL GROWTH. CAN BIND TO DNA
CC AND HAS 5'->3' EXORIBONUCLEASE ACTIVITY.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: TO YEAST RAT1; SOME, TO YEAST KEM1/DST2 AND S.POMBE
CC EXO2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D17752; BAA04601.1;
DR EMBL; Z69240; CAA93235.1;
DR PIR; S43891; S43891.
KW Nuclear protein; Hydrolase; Nuclease; Exonuclease.
FT DOMAIN 268
FT SEQUENCE 991 AA; 112367 MW; 83FA34D93DADAD00 CRC64;

Query Match 42.3%; Score 41; DB 1; Length 991;
Best Local Similarity 41.2%; Pred. No. 56;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGIS 17
   |||||
DB 706 NPTDLNGKKFEGQVGA 722

RESULT 7
FLAV_SYN7
ID FLAV_SYN7 STANDARD; PRT; 169 AA.
AC P10340;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FLAVODOXIN.
GN ISIB.
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86086879; PubMed=3121586;
RA Laudenbach D.E., Reith M.E., Straus N.A.;
RT Isolation, sequence analysis, and transcriptional studies of the
RT flavodoxin gene from Anacystis nidulans R2.;
RL J. Bacteriol. 170:258-265(1988).
RN [2]
RP SEQUENCE OF 1-55, AND X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=83216115; PubMed=6406674;
RA Smith W.W., Patridge K.A., Ludwig M.L., Petsko G.A., Tsernoglou D.,
RA Tanaka M., Yasunobu K.T.;
```

RT "Structure of oxidized flavodoxin from Anacystis nidulans.";
 RL J. Mol. Biol. 165:737-755(1983).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
 RX MEDLINE=20079529; PubMed=10610791;
 RA Drennan C.L., Fattridge K.A., Weber C.H., Metzger A.L., Hoover D.M.,
 RL Ludwig M.L.;
 RT "Refined structures of oxidized flavodoxin from Anacystis nidulans.";
 RL J. Mol. Biol. 294:711-724(1999).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
 RX MEDLINE=20079530; PubMed=10610792;
 RA Hoover D.M., Drennan C.L., Metzger A.L., Osborne C., Weber C.H.,
 RL Fattridge K.A., Ludwig M.L.;
 RT "Comparisons of wild-type and mutant flavodoxins from Anacystis
 nidulans. Structural determinants of the redox potentials.";
 RL J. Mol. Biol. 294:725-743(1999).
 RN [5]
 RP STRUCTURE BY NMR.
 RX MEDLINE=91329335; PubMed=1907844;
 RA Clubb R.T., Thanabal V., Osborne C., Wagner G.;
 RL "1H and 15N resonance assignments of oxidized flavodoxin from
 Anacystis nidulans with 3D NMR.";
 RL Biochemistry 30:7718-7730(1991).
 CC -1- FUNCTION: LOW-POTENTIAL ELECTRON DONOR TO A NUMBER OF REDOX
 CC ENZYMES.
 CC -1- COFACTOR: FMN.
 CC -1- INDUCTION: BY IRON-STRESS.
 CC -1- SIMILARITY: BELONGS TO THE FLAVODOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M19116; AAA22050.1; -;
 DR PIR; A05103; A05103.
 DR PIR; A28670; A28670.
 DR PDB; 1OFV; 31-JAN-94.
 DR PDB; 1CZH; 29-DEC-99.
 DR PDB; 1CZK; 29-DEC-99.
 DR PDB; 1CZL; 29-DEC-99.
 DR PDB; 1CZN; 29-DEC-99.
 DR PDB; 1CZO; 29-DEC-99.
 DR PDB; 1CZR; 29-DEC-99.
 DR PDB; 1CZU; 29-DEC-99.
 DR PDB; 1D03; 29-DEC-99.
 DR PDB; 1D04; 29-DEC-99.
 DR InterPro; IPR001226; Flavodoxin.
 DR Pfam; PF00258; flavodoxin; 1.
 DR PROSITE; PS00201; FLAVODOXIN; 1.
 DR Electron transport; Flavoprotein; FMN; 3D-structure.
 FT INIT_MET 0 0
 FT CONFLICT 54 54 C -> S (IN REF. 2).
 FT CONFLICT 167 168 FG -> GF (IN REF. 2).
 FT STRAND 3 7
 FT HELIX 13 25
 FT TURN 26 26
 FT TURN 28 30
 FT STRAND 31 35
 FT STRAND 36 38
 FT HELIX 41 46
 FT STRAND 49 53
 FT STRAND 56 57
 FT TURN 58 60
 FT STRAND 61 62
 FT HELIX 64 75
 FT TURN 79 80
 FT STRAND 82 88
 FT TURN 91 96

FT TURN 98 99
 FT HELIX 100 111
 FT TURN 112 113
 FT STRAND 115 116
 FT STRAND 120 121
 FT TURN 123 124
 FT TURN 131 132
 FT STRAND 133 134
 FT TURN 135 136
 FT STRAND 137 138
 FT STRAND 141 143
 FT TURN 145 147
 FT HELIX 149 166
 FT TURN 167 168
 SQ SEQUENCE 169 AA; 18646 MW; CF049E12F9C6AALF CRC64;
 Query Match 41.2%; Score 40; DB 1; Length 169;
 Best Local Similarity 53.3%; Pred. No. 13;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 2 PTVALYGLKQDWEGI 16
 || : | : |||||
 Db 55 PTWNVGELQSDWEGI 69
 RESULT 8
 DCOP_BACSU
 ID DCOP_BACSU STANDARD; PRT: 239 AA.
 AC P25971;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE OROTIDINE 5'-PHOSPHATE DECARBOXYLASE (EC 4.1.1.23) (OMP
 DE DECARBOXYLASE) (OMPDCase).
 GN PYRF.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91225016; PubMed=1709162;
 RA Quinn C.L., Stephenson B.T., Switzer R.L.;
 RT "Functional organization and nucleotide sequence of the Bacillus
 subtilis pyrimidine biosynthetic operon.";
 RL J. Biol. Chem. 266:9113-9127(1991).
 CC -1- CATALYTIC ACTIVITY: OROTIDINE-5'-PHOSPHATE = UMP + CO(2).
 CC -1- PATHWAY: SIXTH AND LAST STEP IN THE BIOSYNTHESIS OF PYRIMIDINES.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: BELONGS TO THE OMP DECARBOXYLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M59757; AAA21273.1; -;
 DR EMBL; Z99112; CAB13429.1; -;
 DR PIR; I39845; I39845.
 DR Subtilist; BG10719; pyrF.
 DR InterPro; IPR001754; OMPDecase.
 DR Pfam; PF00215; OMPDecase; 1.
 DR PROSITE; PS00156; OMPDECASE; 1.
 DR Pyrimidine biosynthesis; Lyase; Decarboxylase; Complete proteome.
 FT ACT_SITE 62 62 BY SIMILARITY.
 SQ SEQUENCE 239 AA; 25992 MW; DB1743714ED052E7 CRC64;
 Query Match 41.2%; Score 40; DB 1; Length 239;

CC EMBL; AF213822; AAF23789.1; -
DR InterPro; IPR000728; AIRS_related.
DR Pfam; PF00586; AIRS; 2
DR Purine biosynthesis; Ligase; ATP-binding.
KW NP_BIND 106 117
FT SEQUENCE 734 AA; 77679 MW; B771635E0F66A166 CRC64;
SQ

Query Match 41.2%; Score 40; DB 1; Length 734;
Best Local Similarity 58.3%; Pred. No. 60;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 PTVALYGLKQDW 13
Db 550 PTIGVGLQDW 561
II: |||||

RESULT 12
NKCL MANSE STANDARD; PRT; 1060 AA.
AC Q25479;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE BUMETANIDE-SENSITIVE SODIUM-(POTASSIUM)-CHLORIDE COTRANSPORTER
(NA-K-CL SYMPORTER).
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Spingioidea; Spingidae; Sphinginae; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Malpighian tubules;
RX MEDLINE=96035837; PubMed=7550244;
RA Reagan J.D.;
RT "Molecular cloning of a putative Na(+)-K(+)-2Cl-cotransporter from
the Malpighian tubules of the tobacco hornworm, Manduca sexta.";
RL Insect Biochem. Mol. Biol. 25:875-880(1995).
CC -!- FUNCTION: ELECTRICALLY SILENT TRANSPORTER SYSTEM WHICH IS A
MEDIATOR OF SODIUM AND CHLORIDE REABSORPTION. PLAYS A VITAL ROLE
IN THE REGULATION OF IONIC BALANCE AND CELL VOLUME.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; U17344; AAA75600.1; -
DR InterPro; IPR002293; AA_rel_permease.1.
DR InterPro; IPR002027; Amino acid permease.
DR InterPro; IPR002443; NaKCl transport.
DR Pfam; PF00324; aa_permeases; 1.
DR PRINTS; PR01207; NAKCLTRNSPRT.
KW Transport; Transmembrane; Glycoprotein.
FT DOMAIN 1 122
FT CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 123 143
FT POTENTIAL.
FT DOMAIN 144 153
FT EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 154 174
FT POTENTIAL.
FT DOMAIN 175 197
FT CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 198 218
FT POTENTIAL.
FT DOMAIN 219 249
FT EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 250 270
FT POTENTIAL.
FT DOMAIN 271 275
FT CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 276 296
FT POTENTIAL.
FT DOMAIN 297 331
FT EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 332
FT POTENTIAL.

FT DOMAIN 353 367
FT TRANSSEM 368 388
FT POTENTIAL.
FT DOMAIN 389 431
FT EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 432 452
FT POTENTIAL.
FT DOMAIN 453 497
FT CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 498 518
FT POTENTIAL.
FT DOMAIN 519 562
FT EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 563 583
FT POTENTIAL.
FT DOMAIN 584 642
FT CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 643 663
FT POTENTIAL.
FT DOMAIN 664 881
FT EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 882 902
FT POTENTIAL.
FT DOMAIN 903 1060
FT CYTOPLASMIC (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 396 396
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 404 404
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 419 419
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 816 816
SQ SEQUENCE 1060 AA; 117392 MW; 4C39D1F0B645F5FEF CRC64;

Query Match 41.2%; Score 40; DB 1; Length 1060;
Best Local Similarity 58.3%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 PTVALYGLKQDW 13
Db 719 PNVLMLGYKSDW 730
I: |||||

RESULT 13
DDX8_HUMAN STANDARD; PRT; 1220 AA.
AC Q14562;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ATP-DEPENDENT HELICASE DDX8 (RNA HELICASE HRH1) (DEAH-BOX PROTEIN 8).
GN DDX8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021291; PubMed=7935475;
RA Ono Y., Ono M., Shimura Y.;
RT "Identification of a putative RNA helicase (HRH1), a human homolog of
yeast Prp22.";
RL Mol. Cell. Biol. 14:7611-7620(1994).
RN [2]
RP FUNCTION.
RX MEDLINE=96196415; PubMed=8608946;
RA Ono M., Shimura Y.;
RT "A human RNA helicase-like protein, HRH1, facilitates nuclear export
of spliced mRNA by releasing the RNA from the spliceosome.";
RL Genes Dev. 10:997-1007(1996).
CC -!- FUNCTION: FACILITATES NUCLEAR EXPORT OF SPLICED MRNA BY RELEASING
THE RNA FROM THE SPLICEOSOME.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DOMAIN: THE RS DOMAIN CONFERS A NUCLEAR LOCALIZATION SIGNAL, AND
APPEARS TO FACILITATE THE INTERACTION WITH THE SPLICEOSOME.
CC -!- SIMILARITY: BELONGS TO THE "DEAD" BOX FAMILY HELICASE. DEAH
SUBFAMILY. DDX8/PRP22 ORTHOLOG.
CC -!- SIMILARITY: CONTAINS 1 'SI MOTIF' DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC

```

DR DMBL; D50487; BAA09078.1; -.
DR HSSP; P05035; ISRO.
DR MIM; 600396; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002464; DEAH_ATP_helcse.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR003029; S1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00575; S1; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PS00690; DEAH_ATP_HELICASE; 1.
KW mRNA processing; mRNA splicing; Helicase; ATP-binding;
KW Nuclear protein.
FT DOMAIN 172 175 POLY-LYS
FT DOMAIN 176 228 ARG/SER-RICH (RS DOMAIN)
FT DOMAIN 265 336 S1 MOTIF
FT NP_BIND 588 595 ATP (POTENTIAL)
FT SITE 685 688 DEAH BOX
FT MUTAGEN 594 594 K>E: IN GET; INHIBITION OF PRE-MRNA
      SPlicing AND NUCLEAR EXPORT OF UNSPLICED
      RNA.
FT MUTAGEN 717 717 S->L: IN LAT; INHIBITION OF PRE-MRNA
      SPlicing AND NUCLEAR EXPORT OF UNSPLICED
      RNA.
FT SEQUENCE 1220 AA; 139314 MW; 17C1602A73A0EF24 CRC64;
Query Match 41.2%; Score 40; DB 1; Length 1220;
Best Local Similarity 53.8%; Pred. No. 1e+02;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 4 VALYGLKQDWEGI 16
DB 282 VQLEGLRKRWEGL 294
      | | | | |
      | | | | |
RESULT 14
CARA_METJA
ID CARA_METJA STANDARD; PRT; 354 AA.
AC Q58425;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CARBAMOYL-PHOSPHATE SYNTHASE SMALL CHAIN (EC 6.3.5.5) (CARBAMOYL-
DE PHOSPHATE SYNTHETASE GLUTAMINE CHAIN).
GN CARA OR MJ1019.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
[1]
SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RA jannaschii";
RT Science 273:1058-1073(1996).
RL Science 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: 2 ATP + GLUTAMINE + CO(2) + H(2)O -> 2 ADP +
CC -!- ORTHOPHOSPHATE + GLUTAMATE + CARBAMOYL PHOSPHATE.
CC -!- PATHWAY: INVOLVED IN BOTH ARGinine AND PYrimidine BIOSYNTHESIS.
CC -!- SUBUNIT: COMPOSED OF TWO CHAINS; THE SMALL (OR GLUTAMINE) CHAIN
CC PROMOTES THE HYDROLYSIS OF GLUTAMINE TO AMMONIA, WHICH IS USED
CC BY THE LARGE (OR AMMONIA) CHAIN TO SYNTHESIZE CARBAMOYL PHOSPHATE.
CC
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U67544; AAB99021.1; -.
CC HSSP; P00907; 1JDB.
CC TIGR; M31019; -.
CC InterPro; IPR001317; CPS_GATase.
CC InterPro; IPR002474; CPSase_sm_chain.
CC InterPro; IPR000991; GATase_1.
CC Pfam; PF00986; GATase_sm_chain; 1.
CC Pfam; PF00117; GATase; 1.
CC PRINTS; PR00096; GATASE.
CC PRINTS; PR00099; CPSGATASE.
CC PROSITE; PS00442; GATASE_TYPE_I; FALSE_NEG.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase;
KW Complete proteome.
FT DOMAIN 1 ? CPSASE.
FT DOMAIN ? 354 GLUTAMINE AMIDOTRANSFERASE.
FT ACT_SITE 250 250 GATASE (BY SIMILARITY).
FT SEQUENCE 354 AA; 39920 MW; 4D8F0776CA65F9F CRC64;
Query Match 40.7%; Score 39.5; DB 1; Length 354;
Best Local Similarity 40.0%; Pred. No. 34;
Matches 8; Conservative 4; Mismatches 5; Indels 3; Gaps 1;
QY 2 PTVALYGLKQDW---EGISS 18
DB 56 PLEGNYGVKKDWFSDDGIKA 75
      | | | | |
      | | | | |
RESULT 15
NODA_RHIS3
ID NODA_RHIS3 STANDARD; PRT; 196 AA.
AC P72329;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NODULATION PROTEIN A (EC 2.3.1.-).
GN NODA.
OS Rhizobium sp. (strain N33).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=103798;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=96303537; PubMed=8755627;
RA Cloutier J., Laberge S., Prevost D., Antoun H.;
RT "Sequence and mutational analysis of the common nodBCIJ region of
RT Rhizobium sp. (Oxytropis arctobia) strain N33, a nitrogen-fixing
RT microsymbiont of both arctic and temperate legumes.";
RL Mol. Plant Microbe Interact. 9:523-531(1996).
CC -!- FUNCTION: N-ACYLTRANSFERASE REQUIRED FOR NODULATION. ACTS IN THE
CC PRODUCTION OF A SMALL, HEAT-STABLE COMPOUND (NOD) THAT STIMULATES
CC MITOSIS IN VARIOUS PLANT PROTOPLASTS.
CC -!- SUBCELLULAR LOCATION: CYTOSOL; CYTOSOL.
CC -!- SIMILARITY: BELONGS TO THE NODA FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----

```

```

CC -----
DR EMBL; U53327; AAB16892.1; -.
DR InterPro; IPR003484; Noda.
DR Pfam; PF02474; Noda; 1.
DR PROSITE; PS01349; NODA; 1.
KW transferase; Acyltransferase; Nodulation.
SQ SEQUENCE 196 AA; 21727 MW; CA065D2046F3F061 CRC64;

Query Match      40.2%; Score 39; DB 1; Length 196;
Best Local Similarity 46.2%; Pred. No. 22;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 4 VALYGLKQDWEGI 16
   : |||:: ||:
Db 88 IGLYGVRPDLEGL 100

```

Search completed: March 14, 2002, 09:24:53
Job time: 909 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: March 14, 2002, 09:10:45 ; Search time 56.96 seconds
(without alignments)
24.072 Million cell updates/sec

Title: US-09-765-739a-6

Perfect score: 97

Sequence: 1 NPTVALYGLKQDWEGISS 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	97	100.0	278	JE0216	28k surface antige
2	62	63.9	280	JE0217	28k surface antige
3	58	59.8	276	JE0218	28k surface antige
4	56	57.7	284	I40882	major antigenic pr
5	51	52.6	286	JE0219	28k surface antige
6	48	49.5	133	JE0221	28k surface antige
7	45.5	46.9	122	G64558	hypothetical prote
8	45	46.4	1077	A95533	probable zinc meta
9	44	45.4	162	F82084	probable 2-demethy
10	44	45.4	328	G71838	probable nadh oxid
11	44	45.4	856	G71133	probable alpha-man
12	43.5	44.8	122	A71950	hypothetical prote
13	43	44.3	132	T44301	hypothetical prote
14	43	44.3	298	S75205	hypothetical prote
15	43	44.3	389	T43979	hypothetical prote
16	43	44.3	412	T09313	immediate-early pr
17	42.5	43.8	370	T40131	hypothetical prote
18	42	43.3	178	T29345	hypothetical prote
19	42	43.3	373	G75073	hypothetical prote
20	42	43.3	757	C84120	subtilisin-type pr
21	42	43.3	1023	A47296	thiazide-sensitive
22	41	42.3	168	S18374	flavodoxin - Anaba
23	41	42.3	170	S04600	flavodoxin - Anaba
24	41	42.3	395	H72222	conserved hypothet
25	41	42.3	439	E82426	phosphoglycerate t
26	41	42.3	452	D70768	hypothetical prote
27	41	42.3	480	T15839	hypothetical prote
28	41	42.3	544	S25101	CTP synthase (EC 6
29	41	42.3	918	S04255	regulatory protein

30 41 42.3 991 2 S43891 dna exoribonucleas
31 40.5 41.8 180 2 T16668 hypothetical prote
32 40 41.2 105 2 F83935 hypothetical prote
33 40 41.2 135 2 G83495 hypothetical prote
34 40 41.2 169 2 A42716 antirestriction pr
35 40 41.2 170 1 A28670 flavodoxin [valida
36 40 41.2 239 2 I39845 orotidine-5'-phosp
37 40 41.2 254 2 F69309 ATP-binding protei
38 40 41.2 341 2 S73685 hypothetical prote
39 40 41.2 374 2 H71091 hypothetical prote
40 40 41.2 661 2 S45131 probable membrane
41 40 41.2 774 2 G71308 probable aminopept
42 40 41.2 1060 2 T30823 bumetanide sensiti
43 40 41.2 1122 2 T28130 hypothetical prote
44 40 41.2 1220 2 A56236 probable RNA helic
45 40 41.2 1226 2 T49915 pre-mRNA splicing

ALIGNMENTS

RESULT 1
JE0216
28k surface antigen 3 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0216
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.E
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0216
A:Molecule type: DNA
A:Residues: 1-278 <RED>
A:Cross-references: GB:AF062761

Query Match 100.0%; Score 97; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
DB 60 NPTVALYGLKQDWEGISS 77
|||||

RESULT 2
JE0217
28k surface antigen 4 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.E
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0217
A:Molecule type: DNA
A:Residues: 1-280 <RED>
A:Cross-references: GB:AF062761

Query Match 63.9%; Score 62; DB 2; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.014;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
DB 60 NTTIGVFLKQDWGD 74
| | | | |

RESULT 3
JE0218
28k surface antigen 5 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0218
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0218; MUID:98321180
A:Accession: JE0218
A:Molecule type: DNA
A:Residues: 1-276 <RED>
A:Cross-references: GB:AF062761

Query Match 59.8%; Score 58; DB 2; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.063;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
Db 59 NTTVGVLKQKNDG 73
| | | : | | | : | |
| | | : | | | : | |

RESULT 4
I40882
major antigenic protein - heartwater rickettsia
C:Species: Cowdria ruminantium (heartwater rickettsia)
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40882; S42827
R:van Vliet, A.H.; Jongejans, F.; van Kleef, M.; van der Zelfst, B.A.
Infect. Immun. 62, 1451-1456, 1994
A:Title: Molecular cloning, sequence analysis, and expression of the gene encoding the
A:Reference number: I40882; MUID:94178956
A:Accession: I40882
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-284 <RES>
A:Cross-references: EMBL:X74250; NID:g454266; PIDN:CAA52309.1; PID:g454267
C:Genetics:
A:Gene: map1

Query Match 57.7%; Score 56; DB 2; Length 284;
Best Local Similarity 50.0%; Pred. No. 0.14;
Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEGISS 18
Db 63 TKAVFGLKDKDGVKT 78
| | : | | | : | |
| | : | | | : | |

RESULT 5
JE0219
28k surface antigen 2 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0219
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0219; MUID:98321180
A:Accession: JE0219
A:Molecule type: DNA
A:Residues: 1-286 <RED>
A:Cross-references: GB:AF062761

Query Match 52.6%; Score 51; DB 2; Length 286;
Best Local Similarity 50.0%; Pred. No. 0.92;

Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWE 14
Db 60 NTTVGVEGIEQDWD 73
| | | : | | | : | |
| | | : | | | : | |

RESULT 6
JE0221
28k surface antigen 2 - Ehrlichia canis
C:Species: Ehrlichia canis
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 21-Jul-2000
C:Accession: JE0221
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180
A:Accession: JE0221
A:Molecule type: DNA
A:Residues: 1-133 <RED>
A:Cross-references: GB:AF062762; NID:g3327964; PIDN:AAC26722.1; PID:g3327966

Query Match 49.5%; Score 48; DB 2; Length 133;
Best Local Similarity 55.6%; Pred. No. 1.3;
Matches 10; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 3 TVALYGLKQDWEG--ISS 18
Db 62 TTVVYGLKENWAGDAISS 79
| | : | | | : | | |
| | : | | | : | | |

RESULT 7
G64558
hypothetical protein HP0311 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: G64558
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: G64558
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-122 <TOM>
A:Cross-references: GB:AE000549; GB:AE000511; NID:g2313403; PIDN:AAD07384.1; PID:g231

Query Match 46.9%; Score 45.5; DB 2; Length 122;
Best Local Similarity 56.2%; Pred. No. 3;
Matches 9; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 1 NPTVALYGLK-QDWEG 15
Db 55 NQILAFYGLKINDMOG 70
| | : | | | : | | |
| | : | | | : | | |

RESULT 8
A96533
probable zinc metalloproteinase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A96533
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, C.W.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzic

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: A96533
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1077 <STO>
 A:Cross-references: GB:AE005173; NID:gl0120424; PIDN:AAG13049.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F14J22.13
 A:Map position: 1

Query Match 46.4%; Score 45; DB 2; Length 1077;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 ODWEGISS 18
 |||||
 DB 817 ODWEGISS 824
 |||||
 RESULT 9
 F82084
 probable 2-demethylmenaquinone 2-C-methyltransferase (EC 2.1.1.-) VC2366 [similarity] -
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: F82084
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: A82035; MUID:20406833
 A:Accession: F82084
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-162 <HEI>
 A:Cross-references: GB:AE004307; GB:AE003852; NID:g9656934; PIDN:AAF95509.1; GSPDB:GN001
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VC2366
 A:Map position: 1
 C:Keywords: methyltransferase

Query Match 45.4%; Score 44; DB 2; Length 162;
 Best Local Similarity 46.2%; Pred. No. 7.1;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 QY 4 VALYGLKQDWEGI 16
 :||: |||||
 DB 79 LA1AIKNDWEGV 91
 RESULT 10
 G71838
 Probable nadh oxidoreductase I - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
 C:Accession: G71838
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A:Reference number: A71800; MUID:99120557
 A:Accession: G71838
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-328 <ARN>

A:Cross-references: GB:AE001545; GB:AE001439; NID:g4155776; PIDN:AAD06772.1; PID:g415
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: nuoF
 Query Match 45.4%; Score 44; DB 2; Length 328;
 Best Local Similarity 50.0%; Pred. No. 15;
 Matches 12; Conservative 1; Mismatches 3; Indels 8; Gaps 2;
 QY 1 NPTVALYGLK-QDW-----EGI 16
 ||||| | | | |
 DB 203 NPTVAFYDSKDQEWLLETAFAKEGI 226

RESULT 11
 G71133
 Probable alpha-mannosidase - Pyrococcus horikoshii
 C:Species: Pyrococcus horikoshii
 C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
 C:Accession: G71133
 R:Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oga
 DNA Res. 5, 55-76, 1998
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil-
 A:Reference number: A71000; MUID:98344137
 A:Accession: G71133
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-856 <KAW>
 A:Cross-references: GB:AP000003; NID:g3236130; PIDN:BAA29929.1; PID:dl030872; PID:g3;
 A:Experimental source: strain OT3
 A:Note: this accession replaces an interim accession for a sequence replaced by GenB
 C:Genetics:
 A:Gene: PH0835

Query Match 45.4%; Score 44; DB 2; Length 856;
 Best Local Similarity 63.6%; Pred. No. 41;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 6 LYGLKQDWEGI 16
 ||||: |||||
 DB 309 LYGIEYPWEGI 319

RESULT 12
 A71950
 hypothetical protein jhp0296 - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
 C:Accession: A71950
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric f
 A:Reference number: A71800; MUID:99120557
 A:Accession: A71950
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-122 <ARN>
 A:Cross-references: GB:AE001466; GB:AE001439; NID:g4154813; PIDN:AAD05873.1; PID:g415
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: jhp0296

Query Match 44.8%; Score 43.5; DB 2; Length 122;
 Best Local Similarity 56.2%; Pred. No. 6.4;
 Matches 9; Conservative 2; Mismatches 4; Indels 1; Gaps 1;
 QY 1 NPTVALYGLK-QDWEG 15

Db 55 NQILAFGLKIGDWQG 70
| : | | | | | | : | : |
| : | | | | | | | : | : |

RESULT 13

T44301
Hypothetical protein BH0656 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 31-Dec-2000
C:Accession: T44301; H83731
R:Takami, H.; Nakasone, K.; Ogasawara, N.; Hirama, C.; Nakamura, Y.; Masui, N.; Fuji, F.;
Extremophiles 3, 29-34, 1999
A:Title: Sequencing of three lambda clones from the genome of alkaliphilic Bacillus sp.
A:Reference number: 222745; MUID:99184646
A:Accession: T44301
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-132 <TAK>
A:Cross-references: EMBL:AB011836; NID:g4512345; PIDN:BAA75320.1; PID:g4512355
R:Dominguez, G.; Dambaugh, T.R.; Stamey, F.R.; Dewhurst, S.; Inoue, N.; Pellett, P.E.
J. Virol. 73, 8040-8052, 1999
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
Nucleic Acids Res. 28, 4317-4331, 2000
A:Reference number: A83650; MUID:20263314
A:Accession: H83731
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-132 <STO>
A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA804375.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0656
C:Superfamily: hypothetical protein b2531

Query Match 44.3%; Score 43; DB 2; Length 132;
Best Local Similarity 72.7%; Pred. No. 8.4;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 7 YGLKQDWEGIS 17
| | | | | | | |
Db 65 YRLKRDWEIS 75

RESULT 14

S75205
Hypothetical protein sir2052 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S75205
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S75205
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-298 <KAN>
A:Cross-references: EMBL:D90903; GB:AB001339; NID:g1652127; PIDN:BAAL7119.1; PID:g165219
C:Superfamily: Synechocystis hypothetical protein sir2052

Query Match 44.3%; Score 43; DB 2; Length 298;
Best Local Similarity 42.1%; Pred. No. 20;
Matches 8; Conservative 3; Mismatches 6; Indels 2; Gaps 1;

QY 2 PTVALYGC--LKQDWEGISS 18
| | | | | : | : |
Db 278 PTVIAYGKRISRSWGVQS 296

RESULT 15

T43979
Hypothetical protein U19 [imported] - human herpesvirus 6
C:Species: human herpesvirus 6
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 28-Jul-2000
C:Accession: T43979; T44165
R:Iseigawa, Y.; Mukai, T.; Nakano, K.; Kagawa, M.; Chen, J.; Mori, Y.; Sunagawa, T.; K
J. Virol. 73, 8053-8063, 1999
A:Title: Comparison of the complete DNA sequences of human herpesvirus 6 variants A a
A:Reference number: 222732; MUID:99412319
A:Accession: T43979
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-389 <ISE>
A:Cross-references: EMBL:AB021506; NID:g4995977; PIDN:BAA78240.1; PID:g4996007
R:Dominguez, G.; Dambaugh, T.R.; Stamey, F.R.; Dewhurst, S.; Inoue, N.; Pellett, P.E.
J. Virol. 73, 8040-8052, 1999
A:Title: Human herpesvirus 6B genome sequence: coding content and comparison with hum
A:Reference number: 222734; MUID:99412318
A:Accession: T44165
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-389 <DOM>
A:Cross-references: EMBL:AF157706; PIDN:AAD49631.1
A:Experimental source: strain Z29; variant B
C:Genetics:
A:Gene: U19
C:Superfamily: human herpesvirus 6 hypothetical protein U19

Query Match 44.3%; Score 43; DB 2; Length 389;
Best Local Similarity 33.3%; Pred. No. 26;
Matches 5; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

QY 4 VALYGLKQDWEGISS 18
| : | : | : | : | : |
Db 79 MSFMFGVTEWEGASA 93

Search completed: March 14, 2002, 09:10:46
Job time: 362 sec